| | |
|---|---|
| Chief Investigator | Prof Fiona Lobban, Lancaster University |
| Study Co-ordinator | Dr Tamara Rakic |
| Sponsor | Lancaster University |
| Host NHS Organisation | Berkshire NHS Foundation Trust |
| Funder(s) | NIHR- HS&DR |
| Key Protocol Contributors | Research Team |
| Committees | Independent Study Steering committee chaired by Prof Steven Gillard, City University London |

# Realist evaluation of online mental health communities to improve policy and practice

Mental health problems, including depression, bipolar disorder, anxiety, eating disorders, and psychosis, affect approximately 1 in 6 people in the UK at any one time (>11million people)[1]. People seek support in a wide variety of ways, increasingly through peer support, and by going online[2, 3]. They search for information and join online mental health communities (also called forums). Although widely used, the impacts of using online mental health communities, and mechanisms by which these occur, are poorly understood. Our aim is to improve the uptake, safety and usefulness of online mental health communities. We will do this by developing a programme theory about how online mental health communities work, for who, why, and in what contexts. We will use this theory to develop best practice tools for commissioners, community hosts, moderators, and referring clinicians.

## Background and rationale

The specific aims of the study were developed through an iterative process of reviewing relevant literature, team discussion, expert consultation with stakeholders working in the field, and a series of PPI events.

### *What is the problem being addressed?*

Peer support is "what we do when we recognise our shared experiences of disadvantage, and distress; make an interpersonal connection on that basis; and come together to support and learn from each other" [4]. It can be offered in a range of ways, including individual/group, face to face/ telephone and increasingly online. Online mental health communities are dedicated support platforms aimed at helping individuals to share, discuss and solicit information and support related to mental health[5]. Many are moderated, and some have access to additional online tools and resources. Although people talk about mental health online in a variety of online spaces including Facebook, Twitter, Instagram etc., the focus of this study is on peer communities set up by a healthcare provider, charity, or commercial organisation with the explicit purpose of facilitating peer to peer support for people with mental health difficulties.

Online communities are particularly valued by people with rare or highly stigmatised difficulties, including mental health, and communities traditionally under-served by healthcare services, including men[6], those living in remote areas, or close knit communities wanting anonymity[7]. Online communities offer robust support during pandemics when healthcare services are restricted, and for people in low- and middle-income countries where access to mobile technology is common but healthcare is not. Adults diagnosed with schizophrenia are as likely as adults without any mental health difficulties to form social connections online, despite having fewer offline relationships, lower income, and less internet access[8]. Despite their widespread use, surprisingly little is known about how online communities work and their impacts on members.

Online communities can provide round-the-clock opportunities for members to feel understood, access information, make friends, and use their experiences to help others[9], whilst being able to manage how they present themselves online[10-12]. These communities can function as an important gateway to other offline support as people test out sharing their problems, and are encouraged to access help [13-16]. Naslund et al [10] identify this as a critical point in an individuals' experience that has the potential to challenge stigma and offer support. However, they also highlight the risks of disinformation, dependency,

and sharing of unhelpful strategies. Information shared can be misleading, inaccurate, harmful, or triggering [9, 17]. Hearing about the experiences of others can exacerbate low mood and risk of harms[18], and generate unrealistic expectations and greater anxiety or confusion about one's own condition[19]. Normative and reductionist ideas about health may be predominant, leaving some users feeling misunderstood or even bullied, and people in mental distress may be particularly vulnerable to the negative impacts of this kind of behaviour[20]. Privacy can be a challenge with members worried about how their data might be used by hosts who rely on advertising or data exploitation to offset costs or make a profit[21]. The lack of an evidence-based understanding of how and why online communities impact on mental health outcomes is also a problem for healthcare staff, community hosts and mental health commissioners. Healthcare staff concerns about possible harms of online communities leads to reluctance to recommend specific communities without clearer guidance on criteria by which to make this decision[22]. There are an increasing numbers of communities on offer but little guidance for hosts on how best to design forums, or for commissioners to help them to decide which they should commission. The CQC 2020 community mental health survey of over 17,000 people who received treatment for a mental health condition reported *"Over a third of people (37%) did not receive support in joining a group or taking part in an activity, but would have liked this"[23] .* Improving the design and delivery of online communities is one way to begin to address this, and overcome some of the identified barriers including availability and accessibility for those in rural areas or with physical disabilities.

***Why is this research important?***

This research will produce interdisciplinary in-depth understanding and explanation (theories) of the mechanisms underpinning the positive and negative impacts of online mental health communities. These evidence-based theories will be used to co-design urgently needed good practice guidance and tools. Online knowledge exchange tools will ensure that mental health commissioners, mental health clinicians, and the general population, all have a better understanding of how online mental health communities work, and how to identify those likely to be safe and supportive. Mental health providers wanting to add an online mental health community to their service offer will have commissioning guidance on how to evaluate or compare existing forums using a standardised framework, and/or guidance on how to design their own safe and effective community. Community moderators will be able to access online training in the skills for this complex role, and become part of an ongoing community to share learning, and support their practice. People experiencing mental distress will have guidance on what kinds of forums are likely to lead to what impacts for them, and how this might happen.

The study directly addresses two of the James Lind Alliance top ten priorities for digital mental health; "what are the benefits and risks of delivering mental health care through technology?" and "how can social media be used more effectively to bring people with mental health problems together and help them connect?"[24]. It addresses a key Department of Health and Social Care goal for mental health research 2020-2030 i.e. to improve choice of, and access to, mental health care, treatment and support in hospital and community settings (https://acmedsci.ac.uk/file-download/63608018). The research is particularly timely in the context of the Covid-19 crisis, during which there has been an increase in people seeking help for severe mental difficulties, a decline in the provision of in-person support, and a turn to online communities for mental health support[25, 26].

***Pre-application work***

We reviewed existing literature to ensure our study addressed a clear need, had not already been done, and drew on the most relevant and up to date findings available. We initially searched for relevant literature using PubMed for papers published in the last 20 years, using the search terms "online OR digital" AND "peer OR community OR forum" AND "psych* OR mental OR distress". We identified key papers based on relevance to our topic of interest and added additional relevant papers from reference lists. To search the grey literature we used the same terms in "Google" and examined the top 100 hits. Additional literature was identified during expert consultation with our PPI, Hosts, and Impact Group (detail below). We checked all studies funded in NIHR and clinical trials databases and confirmed that none address the aims of this study. We searched the PROSPERO database to confirm there were no registered protocols of realist informed reviews relevant to understanding the impacts of mental health online communities. Our literature review and stakeholder engagement confirmed the increasing use of online communities, lack of understanding of how online communities "work", and the urgent need to be able to evaluate communities

and better support community moderators. The review also highlighted the rich and diverse literature available to develop our programme theory using realist informed methods as described in Work-stream 1. ***PPI-*** We ran two online PPI events with 12 people in each with lived experience of moderating and/or regularly using online communities. Moderators reported significant challenges in their roles, impacting on their own mental health, and were very keen to receive better training and support in their role. They suggested the tools developed to support moderators could be piloted in a "sandpit" community before being included in final guidance or e-learning tools. Community members identified the need for healthcare staff and commissioners to be more aware of the role online communities can play, in order to enhance appropriate referrals and secure funding, and highlighted the need to understand the role of "super-users" who post a lot, and "observers" who read but do not post. The PPI group did not like the term "lurkers" previously used in the literature to describe this group, so we use the term "observers". All were keen to be part of the PPI Group if the study is funded. We will recruit from this group and more widely to ensure ethnic, gender and ableness diversity in the group as required***.*** We also met as an online group with our online community ***Hosts.*** We clarified what was needed from the hosts and what they would get from being involved. We developed the design, and made specific changes to our design, including establishing a community of practice during Work-stream 1 to give time for this to establish during the study. Key stakeholders were also consulted and have agreed to be part of our ***Impact Group*** to ensure findings directly inform relevant policy and all outputs are designed to maximise impact in the broader context including senior staff (named below) at NHS England, Health Education England, NHSX, ORCHA (orchahealth.com, independent digital health evaluation organisation), Mental Health Commissioning team in Lancashire Integrated Care System (ICS), Peer support charity network led by Mind. We have also liaised with our NIHR NW Coast ARC, Oxford AHSN and the national mental health implementation network who will support our implementation plan (outlined below).

***Relevance to NHS Policy***

The NHS Longterm Plan[27] commits to an increased role for peer support in health and social care. To implement this Health Education England have published a competency framework guiding a national programme currently underway for training peer support worker roles in mental health services [28]. The focus is on individual peer support workers working in NHS services. However, the evidence of effectiveness for this model of peer support is limited and it has been criticised in light of research suggesting peers employed in these roles often end up delivering para-clinical interventions at odds with the underlying principles of peer support [4, 29]. Hence, there is a need to investigate a broader range of peer support approaches, to understand how they work, and what their impacts are to ensure we harness the potential that peer support can offer.

Peer online communities are part of this broader range, and have received much interest during the Covid-19 pandemic due to their ability to offer remote support at scale and relatively low cost. However, there is a lack of evidence to support commissioning this within an evidence-based healthcare system. Evidence gathering is hampered by not understanding the range of impacts (positive and negative) of joining online communities, and what features of the communities determine these impacts. This in turn means there is no framework to even describe these communities, on dimensions likely to be of most importance. Several very recent government policies will impact on how peer communities are commissioned in the future. The UK "Online Safety Bill" due to be published early 2021 [30], and regulated by Ofcom, will require hosts of online mental health communities to protect users from harm, be more transparent about terms and conditions of use and ensure that these are enforced, and respond to complaints. NHSX Digital Technology Assessment Criteria[31] requires all new digital technologies to meet minimum baseline standards around clinical safety, data protection, technical assurance, interoperability, usability and accessibility. These focus on safety through governance of data, rather than on the benefits and harms created through the social interactions within the communities. The National Institute for Health and Care Excellence (NICE), in collaboration with Public Health England have created an evidence standards framework to help providers of digital health technologies understand the kind of evidence needed to create a convincing case to commissioners[32]. Within this framework, online mental health communities would generally be required to meet the standards in Tier B which include: "measures in place to ensure safety in peer-to-peer communication"; and a commitment to ongoing data collection to "show user outcomes (if relevant)" "improving access to care among hard-to-reach populations" and "fostering good relations

between people with protected characteristics". To support this, NICE also offer the Medtech Early Technical Assessment (META) Tool to help medical technology companies plan evidence generation[33]. Whilst these tools are very useful for some digital technologies such as behaviour change apps, they rely on the technologies having clearly specified measurable impacts, and a clear programme theory about how they work. This is currently lacking for online mental health communities and therefore meeting the required evidence standards will be very challenging. This research aims to address this gap by developing and testing a programme theory to explain who online mental health communities "work", for whom, and under what circumstances. This theory will inform design principles and evaluation guidance that can enhance the evidence standards framework and the META tool, and help commissioners, and hosts to support safe and effective online mental health communities.

***How does the existing literature support this proposal?***

Current research suggests a wide range of impacts from using online mental health communities, but with little understanding of why they vary across individuals, or across different communities. Studies have shown positive outcomes associated with using online communities, including reduction in depression[33] suicidal thoughts[34], and social isolation[7], whilst others have reported negative impacts on suicidal ideation[35] mood[36], body image and disordered eating patterns[37].

Particular features of communities such as size and heterogeneity, target population, ease of access and level of friction, how individual identities are managed (user identification vs anonymity), community rules, and incentives for use and for pro-social behaviour have all been hypothesised as features that might account for this variation in impacts [38-41]. It has been suggested that the ways in which communities broaden access to information, or facilitate sharing of deeper analysis around personal experiences, may be experienced more positively for people who cope with difficulties by seeking deeper understanding, and may be less helpful to those with more avoidant coping styles[17].

The style of moderation is also likely to be important, specifically how moderators are recruited, trained and supported, moderating rules and how they are developed, temporality, transparency and opportunities for redress to moderation decisions. Understanding how these contextual features influence the culture and function of the communities is crucial. This has highlighted the need for research to go beyond the study of behaviour in individual communities and investigate what aspects of design, moderation, and culture differ between communities, which facilitate better outcomes, and how [38, 42].

Research to date has drawn on a wide range of methods, but all have limitations, which impedes firm conclusions when they are used in isolation. These include quantitative surveys, qualitative interviews, computational discourse analysis, and social network analysis. However, many of these studies are cross sectional and descriptive, or tend to only evaluate individual online communities. Randomised controlled trials have little ecological validity as participants choosing to engage in online communities, are already accessing them, and those that are not, are difficult to recruit and retain[33, 43]. Recent investigations using computational linguistic methods have indicated positive changes in emotion over time for those using online mental health communities[44], and developed methods to identify potential mechanisms [5, 45], but these studies are generally limited to a single online community. Our research will move the research field from single community/single methods studies, to triangulate findings from a range of methods to test comprehensive programme theories across different online communities, sampled for contextual diversity. The programme theories will inform co-design of best practice guidance for community hosts and moderators.

***Community Moderators*** play a crucial role in increasing engagement, encouraging discussion of more negative emotions, managing difficult behaviours, building trust, keeping conversations on topic[46], and directing the focus and values of the online community[47]. Moderators may be trained health professionals, people with lived experience, or volunteers (often students), and are often drawn to the role because of their own lived experiences of mental health difficulties. Community members perceive moderator information and advice to be more reliable than that of other members, and value having someone to whom they can refer concerns about other members[7]. Moderators may also play a role in reducing fear of negative evaluation, or fear of causing harm, both of which have been expressed by community members who observe but do not post [17]

However the moderator role is highly variable, ranging from identifying and removing risk related materials, through to that of a highly skilled online counsellor, directing the focus and values of the online community[47]. Some communities require all posts to be pre-moderated before being posted, whereas

others rely on the community to decide what should be flagged / removed. Moderators can also be experienced as overly restrictive, controlling, and limiting the development of peer-to-peer engagement [9, 18, 46]. Moderators often benefit personally from the opportunity to directly support community members, and gain valuable insights into the lived experience of mental health difficulties, but may also suffer high levels of stress and burnout[48]. Communities may host members who are highly distressed, or suicidal, but their level of training and support may not equip moderators to cope with the emotional labour involved in this kind of role. In our pre-application PPI consultation, one moderator described how the sense of responsibility she felt in moderating an online community linked to a large University Hospitals Trust had exacerbated her own mental health difficulties to the extent that she was admitted to hospital. She reported having minimal support from a supervisor with no experience of managing online communities, and as a volunteer, she was not eligible for any employee health or wellbeing services. Understanding how best to recruit, train and support moderators, how moderator behaviour is received and responded to by community members, and how this impacts on member outcomes, are all key to informing good practice guidance and training for moderators.

***Relevant research currently underway***

Relevant research spans many disciplines. Our team includes experts in each of these, all keen to bring them together in a truly interdisciplinary way.

***Digital mental health-*** Much of the current research in digital mental health (including by our team) focuses on progressing mobile apps from tools to provide information, monitor mood, or activity, to tools that can pick up personalised early signs of relapse and provide "just in time interventions" in context, to try to prevent relapse in people with long-term relapsing conditions such as psychosis e.g. ACTISSIST https://sites.manchester.ac.uk/actissist/ . Whilst this work has much value, it frames mental health experiences as symptoms to be prevented, and locates the locus of control within the individual. This is in contrast to online mental health communities, which are more closely aligned to a peer support model, in which the focus is on the "whole of life", unusual experiences are normalised as responses to difficult life circumstances, and even valued, and change comes through relationships and community support[49, 50].

***Use of AI -*** Digital technology being used to support online communities is developing at pace. As more and more people are using online communities, the task of moderating content is also escalating. For communities working with vulnerable populations, and where posts are all pre-moderated, this is a particularly significant challenge. Artificial intelligence is increasingly being used to automate content moderation, and to develop chatbots to respond to questions. Currently the technology can identify key words, and generate standard responses, but is unable to understand or manage the nuances of human interaction, or replicate the skill of experienced moderators. Ongoing research by colleagues in Nottingham (including members of our own team (Rawsthorne), and in Brighton (e.g. [51]) is trying to integrate human moderation features into deep learning classification models (such as BERT) with ensemble methods that can detect mental health related topic words, and improve automated moderation to reduce the burden of moderation task and improve moderators skills. Technological advances in Natural Language Processing also try to "improve" peer responses e.g. PARTNER that can rewrite posts to make them more empathic[52]. Increasingly members will be able to personalise what they see by setting filters for posts with specific content that they want to avoid. However, peer support and moderation are highly complex tasks, and there is a danger that without a deeper understanding of how they "work", these technological advances will be at best missing the point, and at worse, working against the important underlying causal mechanisms of change in peer support. To develop this work further requires a more clearly articulated theory about how different online communities "work" and evolve over time, and the role of moderators in negotiating this process i.e. moving from a "logic of content" to a "logic of care" [53]. This theory will draw on offline work done by our team e.g.[54] and be relevant to online peer support communities as they evolve, even as the technology by which they are delivered changes.

***Balance between safety and protection and therapeutic benefit -*** The UK Online Safety Bill reflects concerns across society about the potential for online spaces to do harm. People with mental health difficulties may be particularly vulnerable to such harm, and therefore there is a desire to regulate and manage online communities to protect people. However, this may have the unintended consequence

of increasing stigma and isolation for people who are struggling with particular issues, such as suicidal ideation. There is much interesting work currently underway to understand how best to balance the need to talk about different issues, and potential for this to be triggering for others, such as that by the Orygen mental health group in Australia (e.g. Affinity Study [17]).

***Increasing access to healthcare -*** Two thirds of people in need of help for mental health difficulties are unable or unwilling to access it through conventional face to face services[1]. Online mental health communities have the potential to engage traditionally underserved populations, may serve as a gateway into mental health services [15], and have the potential to spread effective healthcare information and interventions at scale and speed. Once in online communities, patterns of behaviour differ widely, with approximately 90% of members never posting and between one third and a half of posts being made by 1% of members termed "super-users"[55]. Current research (including our own) uses posting behaviour and metadata to understand the demographic characteristics and people in online communities[56], and social networks analysis to understand how information and support is shared across them. Techniques to do this are becoming more sophisticated, drawing on more features of behaviour e.g. [55, 57]. However, without understanding why different people engage in such different ways, and what the impacts are of doing so; it is difficult to know how to intervene to broaden access and ensure current inequalities in access to mental health support are not replicated online.

## What is the research question / aims and objectives?

### *Aims*

To develop a programme theory to understand the underlying mechanisms by which online mental health communities impact on people's mental health and wellbeing.

To use this programme theory to develop best practice tools to improve uptake, safety and usefulness of online communities.

### *Research Questions*

**RQ1**. **What are the impacts of using online mental health communities for people experiencing mental health difficulties? How, why, in what contexts, and for whom are these impacts generated?**

Objectives

1a. Develop theories of the underlying generative mechanisms by which, and contexts within which, online mental health communities impact on mental health and wellbeing outcomes for members.

1b. Test and refine the theories in case studies of online mental health communities.

**RQ2. What are the roles of community moderators and how do they impact online mental health communities?**

Objectives

2a. Develop a theory to understand the roles of online mental health community moderators, how they vary in different contexts, and how they impact communities.

2b. Test and refine the theories in case studies of online mental health communities.

**RQ3. How can evidence-based theories of online mental health communities be used to inform best practice guidance and support?**

Objective

3a. Co-produce best practice tools, training and implementation plan to optimise the design and delivery of online mental health communities.

## Project Plan / Methods

**Design -** A realist informed mixed methods evaluation, including multiple case studies[58] to co-design theory informed best practice guidance tools for online mental health communities. The cases will be peer online communities supporting mental health. Mixed-methods, using existing data (community posts) and new data, will include:

### *Work-stream 1- Theory building and hypothesis generation*

a) a realist informed review of existing literature and stakeholder theories

***Work-stream 2- Theory testing and refinement***
b) quantitative surveys of community members to assess impacts of being involved in the community
c) qualitative interviews with community members, moderators, hosts and commissioners to understand perceptions of how communities work across a range of different contexts
d) Corpus-based Discourse Analysis and Natural Language Processing of community posts to test hypothesised mechanisms across a range of contexts using "real time" contextualised data.

***Work-stream 3 - Impact***
e) Co-design of best practice tools and implementation strategy

Realist philosophy asserts that complex social interventions in health have intended and unintended impacts (outcomes- O) through the way in which people respond to the resources offered through the program (mechanism - M). These are both triggered by the presence or absence of specific elements in the environment in which the program is delivered (context- C). Understanding what it is about a particular program or intervention that achieves particular outcomes, and why the program results in different outcomes for different people, and in different contexts (CMO configurations) requires in-depth theory building and testing [59, 60]. The realist approach has been chosen because it most readily addresses our research questions.

**Theoretical framework -** We will develop a programme theory to explain the impacts of online mental health communities, for whom, why, and under what circumstances. To do this, we will be guided by existing theories, related to how peer support changes happen in mental health at an individual (micro) level e.g. personal recovery[61], health belief model[62]; an interpersonal (meso) level e.g. peer support[49, 63, 64], social learning theory [65]); and a community organisational (macro) level e.g. sense of community[66], emotional contagion theory[67]). Following the advice of Shearn et. al. [68] and Flynn et. al.[69], we will use these theories to guide the development of our initial CMO hypotheses in work-stream 1, which will be further tested and refined in work-stream 2. Our final programme theories will be used to inform the development of best practice tools in work-stream 3, and may have relevance (though will need further refining in context) for understanding how online communities work in other contexts such as other health conditions, and offline peer support.

**PPI** - Our co-app team includes two people with lived experience of mental health difficulties, and using online communities to manage these. They will be part of the project management team guiding the design and delivery of the project. They are both very experienced in drawing on their lived experience of mental health difficulties to inform research, with Rawsthorne having additional expertise in evaluating online mental health communities [70] using AI to understand human interaction e.g. [54] and in co-design methods[71].
Our three Expert Groups consist of: people with lived experience of mental health difficulties, and moderators (*PPI Group*); hosts from our participating community cases (*Hosts Group*); and key stakeholders who can ensure the outputs are fit for purpose and have a direct pathway to impact (*Impact Group*). Some members across all of these groups have lived experience.
The PPI and Hosts Groups will meet bimonthly (some separately and some together) and, guided by UK standards for public involvement[72, 73] and co-production[74], will be supported to play a key role in developing our theory-informed best practice tools. In Work-stream 1 they will challenge and refine developing hypotheses. In Work-stream 2 they will inform the gathering, integration and interpretation of evidence to consolidate programme theories of how online mental health communities work and the role of moderators. In Work-stream 3 a Co-design subgroup will co-design the theory informed practice tools and implementation strategy.
An independent study *Steering Committee* (including PPI) will oversee the conduct of the study.
Training to support a working understanding of all of the research methods drawn on in this study will be offered in our whole team training event in Work-stream 1. Individual needs for training and support will be regularly reviewed and provided from across the team.

**Setting -** Our research takes place online. Our cases are online mental health communities hosted by NHS funded healthcare providers, mental health charities, or commercial organisations with the explicit purpose of facilitating peer to peer support for people with mental health difficulties.

**Our Hosts.**. All our host communities will be dedicated support platforms aimed at helping individuals to share, discuss and solicit information and support; related to mental health; hosted by healthcare provider, charity, or commercial organisation. They have been purposively sampled for diversity across host, target population, design (including level and nature of moderation, and whether or not they require registered login), and size of population.

Communities recruited to date include; commercial hosts commissioned by the NHS, and Kooth,; openly accessible Reddit (subReddit Mental Health UK); NHS Trust hosted Support Hope and Recovery Online Network SHaRON; and a mental health charity, Bipolar UK. We have the option to include up to six community cases during WS2 if our theory suggests that we need to test our theory in a community with specific characteristics. Where possible we will aim to recruit a more recently established online group, or one representing marginalised or under-served communities.

To facilitate our ethical framework, we have classified the online communities into the following level. Communities in each level have different requirements for consent to use the community posts, based on the design of the communities. This structure is also explained in the flowchaRt and in the data management plan.

Level 1 a posts are publically available without requiring a login and/ or consent for posts to be used for research by other partner organisations has been freely given.
Level 1 b –Consent for research is given, but is tied to terms and conditions of using the site and so is not freely given. Details of the study will be shared with all users, and we will only analyse posts prospectively from telling people about the study and will offer everyone the option to opt out.
Level 1c – consent is not given for research, OR data is linked to health/ social care records and therefore there is a higher expectation of confidentiality. Details of the study will be shared with all users and participants will be required to opt into the study (see PIS and consent form).

Level 2 – Recruit individual community members to survey or interviews, but no analysis of posts.

Level 3 – No community established, but organisation are keen to learn how to overcome barriers to doing so. Hosts are involved in Expert groups but no data collected from community members

Level 4 – Posts only. We will also test our evolving programme theories in archived datasets of communities which have been shut down because of concerns that they were harmful to users. These focus on (1) self harm (subReddits r/SelfHarmReddit r/selfharmpics/ r/SelfHarmCommunity) and (2) disordered eating or proana (subreddits r/EDAnnonymous and r/proED). Whilst it is not possible to survey or interview users of closed communities, our programme theories about how peer support in online communities works, and the role of moderators, can be tested in an archive of the community posts which can be accessed using Pushshift archives http://files.pushshift.io/reddit/ for research purposes.

***Kooth*** https://www.koothplc.com/ is a public limited company commissioned by NHS, Local Authorities, charities, and businesses to support people experiencing a wide range of emotional and mental health difficulties. Users have to register and all posts are pre-moderated. The largest community is aimed at children and young people (10,000 users per month) but has a growing (1, 200 new adult users per month) population of adults.
***Reddit (subR Mental Health UK)*** https://www.reddit.com/r/MentalHealthUK/ is a community on the Reddit platform dedicated to providing support and resources aimed mainly at people in the UK dealing with mental health issues. It is open to access but there are clear rules to guide behaviour. The community has 4,500 members, is increasing steadily and is moderated by a volunteer.
***SHaRON*** https://www.sharon.nhs.uk/ (Support Hope and Recovery Online Network) is designed and hosted by Berkshire Healthcare trust. It was originally set up to support people with difficulties around eating, and is now being rolled out across a range of mental and physical health services across the NHS

and education sites. Users are restricted to those referred by health care staff. Trained clinical staff and volunteers moderate the communities. SHaRON has over 3500 registered users.
***Bipolar UK*** https://www.bipolaruk.org/ecommunity is a national charity supporting people with bipolar disorder. The e-community is growing and currently has over 8000 members who are adults with bipolar disorder. It is moderated by a small team of people with lived experience.

**Project Set-up**
Prior to funding we will appoint our research staff, prepare ethics, set up our Steering Group, and convene our PPI, Host and Impact Groups. We will ensure we have our data systems set up, and buy licences for Miro software. We will prepare our whole team training, and liaise with our learning technicians at Lancaster University library regarding work on our realist synthesis.

Our work-streams are timed to be as efficient as possible, and they overlap where it is possible to conduct tasks in parallel.

**Work-stream 1- Theory Building (1-18 months)**

Following established realist methodology[75-77], we will develop our initial candidate theories of how online mental health communities "work" (**RQ1a**) by drawing on relevant "middle-range" theories, reviewing relevant literature (including grey literature and intervention protocols); and conducting in-depth interviews with key stakeholders including a broad range of academics, clinicians, moderators with expertise in online mental health communities. The importance of community moderation on outcome is well established [7, 9, 46, 47], but not well understood. Therefore to answer **RQ2a** we include a nested review question focussing specifically on understanding the role of the moderator and how this works in different contexts.
We will use the framing of Context-Mechanism-Outcomes (CMO) configurations to guide all stages of the process, from theory construction to the development of data collection protocols and data analysis. Outcomes will include all positive and negative impacts on community member emotional wellbeing or functioning e.g. anxiety, mood, personal recovery, quality of life, hope, stigma, mental health literacy, and help seeking. Mechanisms will be defined as the responses of the community members to the community resources (including moderation) e.g. safety, toxicity, empathy, normalisation, bullying, reframing etc. Context will be explored at several levels and include any features of the community, or backdrop against which it is being delivered, or the individual members, that may influence these mechanisms e.g. hosting, design, size, target population, anonymity etc.

We will first develop an initial framework for candidate CMOs based on existing middle range theories, and drawing on the expertise across our team (including co-apps and Expert Groups). We will select middle range theories relevant to each level of analysis: individual (micro-); interpersonal (meso-) and organisational (macro-). Selection will be guided by the extent to which each theory addresses our research aims, and their compatibility with the realist approach and underlying philosophy i.e. to what extent they help to identify generative causal explanations for key outcomes.
Next, and with additional support from our experienced systematic review team at Lancaster University Library, we will develop a strategy to ***review existing literature*** (relevant databases such as Medline, PsychInfo, EMBASE etc.) and additional grey literature (e.g. Google) that reports on the contexts in which online mental health communities are offered; the intended / proposed / measured outcomes; and explicit / implicit theories about how they "work". The ***search strategy*** will be systematic and clearly documented, but will develop iteratively to include more diverse data sources that can shed light on generative causal pathways to outcomes. ***Data sources*** will include primary data, reviews, policy documents, training manuals, theoretical commentaries, author interpretations, and unpublished reports that give a rationale for offering online mental health communities, describe or evaluate online mental health communities, or detail the lived experiences of members /moderators / hosts / commissioners. ***Screening*** (initially by title and abstract, with 10% checked by a second reviewer) will be based on functional criteria of relevance, rigour and richness i.e. whether the document tells us anything about the mechanism and contexts that we are interested in, rather than methodological quality. ***Data extraction*** will focus on extracting explanatory accounts and mapping where this data is coming from (details of the context, nature of the intervention etc.). ***Analysis*** will be done using retroductive reasoning i.e. working back from the data to identify the context –dependent mechanisms underlying the impacts described in the data that can help us to

understand what is happening in online communities. We will work across individual documents to remove duplication and to integrate and consolidate explanatory accounts. Particular attention will be paid to accounts of potential negative impacts ("dark logic models"[78]), and those that are unexpected, contradictory, or challenge our initial rough programme theories. Our review will be iterative and flexible to ensure we focus our resources on the data best able to help build our theory.
In parallel to the literature review, we will interview up to 20 key expert stakeholders. These stakeholders will include academics, clinicians, commissioners, policy makers, and moderators with broad expertise in online mental health communities. We will sample across willingness to support, including some clinicians who do not refer/support, to understand their concerns. We will include social prescribers within our stakeholder sample. ***Sampling*** will be done through snowballing techniques, initially drawing on the networks of our broad research team, members of our Impact Group, and key individuals identified in our literature searching. ***Data collection*** will be done flexibly (to maximise participation) using individual/ group interviews, online / face to face /telephone, and using topic guides designed to elicit CMO explanatory theories. Data will be recorded and transcribed. ***Data analysis*** will be retroductive i.e. working back from participants descriptions of their experiences, to identify the mechanisms underlying these experiences that might account for what is happening in online communities. Retroductive reasoning allows us to move beyond deductive analysis, and develop new theories about the underlying mechanisms of why a particular outcome happened. To facilitate this process we will draw on several strategies including: counterfactual thinking (could X have happened without Y?); social and thought experiments (imagine a hypothetical world in which the outcomes was different – what would have changed?); studying pathological / extreme cases (detailed analysis of people with very positive or negative outcomes to see how the context was different); and comparative case studies (comparing our community cases to understand how and why outcome differences are determined by structural and contextual differences).

As these interviews will be being conducted contemporaneously with the review we will iteratively feed in the emerging findings into the evolving CMO framework. Data sources will be managed in NVIVO.
We will articulate a theory of how peer online communities impact community members. This process will employ iterative refinement, starting with an array of potential outcomes, multiple possible mechanisms, and an initially broad context. Through iterative exploration of expert opinion and literature, and structured collaborative discussion, we will work with our Expert Groups (PPI and Host) through prioritisation and consensus building, to ensure our final explanatory framework is comprehensible and useful for the design of policy and practice tools in WS3.

The first step of our review will be to define the scope. We will develop a shared glossary to define key concepts, to be evolved throughout the project. We will agree the review scope by defining features of mental health online communities to determine what will be included. For example; online spaces to facilitate peer to peer sharing; aimed at young people and adults over 16; set up by a healthcare provider, charity, or commercial organisation with the explicit purpose of facilitating peer to peer support for people with mental health difficulties.

We will then work with our Expert Groups to formulate our rough preliminary theories. We will ask them to describe the offer within the communities they use / host; identify what they aim to achieve in using / hosting; reflect on how they think these aims are realised (or not) and for who, and how the circumstances they operate in impact on these mechanisms.

Based on these, we will create a written narrative and flow-diagram of candidate theories. The Expert Groups will be presented with the candidate lists and asked to prioritise key outcomes. Prioritisation may include focussing on more proximal outcomes, such as emotional support, learning, sense of belonging, reduced stigma etc., rather than more distal outcomes such as personal recovery or quality of life to further sharpen the focus.

We will then start to articulate specific CMO configurations related to prioritised outcomes, drawing on relevant middle range theories relevant to our understanding of underlying mechanisms. Selection will be guided by how much each theory addresses our research aims, the prioritised outcomes, and underlying philosophy. We will consider theories at an individual (micro) level e.g. health belief model [62]; an interpersonal (meso) level e.g. peer

support[49, 63, 64], or social learning theory [65]); and a community organisational (macro) level e.g. sense of community[66], emotional contagion theory[67]).

With our prioritised outcomes, and hypothesised mechanisms, we will then systematically search the literature for other relevant research. Our clearly defined concepts will guide development of the search strategy in collaboration with our University Librarian.

We anticipate identifying contradictory explanatory accounts, which can be directly tested in W2.

The evolving CMO framework will be shared at regular intervals with the study team, PPI, Host and Impact Groups in written and diagrammatic form for ongoing refinement, integration and prioritisation of our initial programme theories to take forward for testing in Work-stream 2. This prioritisation process is crucial in ensuring we identify a manageable set of the most relevant CMO configurations to take forward into Work-stream 2.

Our protocol will be published on PROSPERO and the findings reported in accordance with RAMESES publication guidelines[79].

**Work-stream 2- Theory Testing & Refinement (6-30 months)**

In Work-stream 2, our hypotheses about 1) how online mental health communities work for whom, how and in what context (**RQ1b**), and 2) the role of community moderators (**RQ2b**), will be tested in our communities, sampled for diversity in host, design (including moderation), size and population. Further communities will be recruited if required following theory development, and a sub-analysis will be carried out on community posts in online mental health communities closed due to reported harms.
A ***longitudinal realist evaluation using multiple mixed methods case studies design*** with novel triangulation of survey, interview, and corpus linguistic methods will be used within and across communities to test our realist hypotheses.

***Detailed description of cases***
We will first develop a case report that describes each case individually in detail, including key contextual features (based on aggregate data provided by the host) as they are identified in Work-stream 1 as influencing user experiences e.g. design of community (including access to additional online tools), ease of access and level of friction, size and heterogeneity of user population, how individual identities are managed (user identification vs anonymity), community rules, and incentives for use and for pro-social behaviour. We will describe moderation strategies in each case including how moderators are recruited, trained and supported, moderating rules and how they are developed, temporality, transparency and opportunities for redress to moderation decisions.
Communities participating at level 2-4 will be described using the same data where available.

The data collection protocol for Work-stream 2 will evolve during Work-stream 1 to ensure it directly tests the updated version of the programme theory; examples given here of contexts, mechanisms, and outcomes are illustrative only (informed by our pre-application work and collaboration). Consistent with optimal case study design, and realist approaches, we will take a mixed-method approach to explore relationships between quantitative measures of outcomes, contextual factors and some mediators, but in-depth qualitative exploration of processes underpinning the deeper causal relationships between these variables will ensure that demi-regularities between variables across the sample are not in themselves assumed to be causal.

***Surveys assessing impacts within and across cases***
New and existing member surveys will be conducted to quantitatively assess impacts of being a community member, and evaluate potential mediators and moderators. New members and existing members will be assessed with three waves of data collection at baseline, six weeks, and 12 weeks later. The choice of these intervals is based on times to clinically important change observed in previous online peer support

studies [80, 81]. Two iterations of the survey will allow for additional adaptation of measures should new hypotheses arise needing evaluation.
Participants will be ***recruited*** over a 9 month time period (plus 3 month follow up) across all community cases in parallel and will include members who passively view but rarely or never actively post to the forums (observers. Participants will be offered an opportunity to participate in the survey via a link to the study website. Invitations to the survey will be shared by our hosts on their forums. These will summarise the nature and purpose of the survey, as well as providing a link to the study website. On following that link, potential participants will see eligibility criteria, a participant information sheet (PIS) and consent form. Those participants agreeing to all consents and inclusion conditions (UK residents aged 16 years and over who have visited the forum at least once) will be sent a link via the email that they provide, which will take them to the survey pages, which are hosted on Lancaster University's REDCap or Qualtrics platform. After completing the survey which we anticipate will take no more than 20 minutes, participants will be asked to provide basic demographic information, including a partial postcode. The partial postcode will allow us to locate which Lower-layer Super Output Area (LSOA) the individual resides in. LSOAs are geographical units with populations of around 1,500. Many rich sources of data are collected at the LSOA level, for example the Indices of Multiple Deprivation, which includes such information such as the accessibility of GP services, the rate of crime, housing quality, and joblessness in the LSOA. Having access to this data will allow us to reduce the bias in our sample due to potential environmental factors in the causes of mental health differences. The downside of obtaining partial postcode data is that it will enhance the statistical disclosure risk of the data. We will reduce this risk by converting the postcode information into the associated LSOA-level variables for datasets that may be made available for deposit in research data archives.
Participants will be sent 6-week and 12-week invitation email, additionally we will send up to two additional email reminders for each survey, in case participants do not complete it after the initial invitation. We will also ask participants if they are willing to be contacted for an in-depth interview. We will further ask respondents for permission to obtain and use their behaviour on the online forum, e.g. how often they log in to the forum, how often they make posts to the forum, the content of their posts. The participants will indicate their willingness to allow us this access by providing the username (or email address for some forums) that they use on the forum. We will adhere to British Psychological Society guidance (112) for taking informed consent online including: taking a record of valid consent; check boxes relating to specific consent statements; limiting the number of consent items; and ensuring participants are fully informed of study procedures, risks, confidentiality and right to withdraw.
Observers (community members who do not post) will be approached in the same way as other potential participants. Observers are not merely passive members and many actively read posts, and benefit from community support/ information..
To incentivise recruitment and maximise participant retention for survey waves after the first one, we will offer rewards and send email reminders to participants. We will offer a fixed reward (a £10 voucher) for each completed wave of the survey, i.e. potentially £30 for completing all three waves. The use of incentives and email reminders have been shown to benefit participant recruitment and retention in similar online surveys [119, 120].

The survey design in WS2 will be informed by WS1 to test the updated programme theory. This updated programme theory will be informed by extensive collaboration with our PPI group. The survey team will work with the PPI group to optimise the operational aspects of the survey and the selection of relevant measures to maximise access, acceptability and relevance. Because of the iterative and collaborative nature of the work, we may update the survey version over the course of the recruitment window. A large proportion of the survey content will be common across versions, but where we find that important concepts identified in WS1 have not been included in the original survey version, we will endeavour to include them in the updated survey versions. We have planned for up to two such iterative updates, to take place potentially around four months and eight months after the initial survey roll-out. To maintain the integrity of the longitudinal data we collect over the three survey waves, each individual participant will be presented with the same survey version across all three longitudinal waves.
We will initially roll out the survey on a limited number of forums (two) and a limited number of participants (100 per forum).

***Survey Measures*** will be guided by the developing theories so those listed here are illustrative only. Where established measures are available of hypothesised positive and negative impacts on community member emotional wellbeing or functioning these will be used (e.g. GAD-7 for anxiety [82], PHQ-9 for mood [83]), and where not, questions for each relevant variable will be created. Established scales will be amended as required to ensure we do not elevate participant risks in ways not ameliorated by our protocol. For example, we will not use the final question in the PHQ9, which asks about self-harm and suicide. Respondents will additionally be asked to reflect about perceived change during their engagement with the community. A widely used measure to assess this is the Global Rating of Change (GRC) which has been shown to represent more accurate descriptions of mental state changes than symptom measures [84]. Context will be measured at several levels and include any features of the community, or backdrop against which it is being delivered, or the individual members, that may influence these mechanisms e.g. hosting, design, size, target population, anonymity etc. Potential impact of use of other health resources is relevant to this survey, as with any evaluation of an intervention. We will collect information on use of other services using an adapted Client Service Receipt Inventory (Lobban et al 2020). We will collect information on participant use of healthcare services, personal social services and medicines. We will conduct sensitivity analyses to evaluate the robustness of our models with respect to variation in use of other services. Survey data collected in level 1 cases, will be linked to user names, only with participant consent, to enable level of community engagement to be assessed directly within the community. This linkage will maximise the extent to which the project will benefit from triangulation. This linkage will only be with participant consent, for the period of the research and not available to moderators or other forum users to ensure user anonymity. All community members will be eligible to take part in the survey, and we can compare demographic and usage data of the sample of those who do, with the whole community sample (where this is available). Measuring at multiple time points allows us to describe patterns of change over time for different participants, and compare these across contexts and between cases. For example, we expect less change and more resilience in established users compared to new ones, but without measuring outcomes at multiple time points for all users we have no way of testing this. We use three time points as the minimum required to estimate longitudinal mediation models, which are much less prone to bias than cross-sectional mediation models. Mediation models allow us to evaluate hypothesised mechanisms by which contextual differences bring about changes in outcomes.

***Analysis of survey data within cases*** will include the detailed description of the community sample, and summary statistics on mental health outcomes.

***Analysis of data across cases*** will account for the hierarchal nature of the data. Users interact with other users and moderators by posting messages to the forum. This activity is overseen by the moderators. The messages posted on a given occasion are therefore at the bottom (level 1) of this hierarchy, with users above them (level 2), all nested in individual community cases (level 3). We will analyse these multilevel data using generalized mixed models and generalized structural equation models using the Mplus (v8.6) software package [85, 86]. We used the simulation results of Pan et al. (2018) to evaluate what ***sample size*** would be required to have statistical power of at least 0.8 to detect small ($a*b = 0.02$) and medium ($a*b=0.15$) sized longitudinal mediation effects across three waves of data. We assumed that the Intra-Class Correlations (ICC), i.e. the proportion of variance in the mediator and outcome variables that could be attributed to level 2 and above, would be a relatively unfavourable 0.6. We further assumed that we would use maximum likelihood by bootstrap estimation for our models, and significance level of $p = 0.05$. With these assumptions we would need $N = 385$ participants to detect a small indirect effect. Assuming 20% attrition per wave would increase this number to 602, which is highly feasible across all of our communities. We will attempt to minimise drop-out by working with our PPI group to ensure the survey is presented in a simple, attractive format with no unnecessary questions. We will use prompts to maximise retention to the survey, successfully employed in previous research (Lobban et al 2020). Second, we will adjust our models to account for potentially non-ignorable drop-out using up-to-date statistical methods, such as full-information maximum likelihood estimation and/or multiple imputation (Schmidt & Woll 2017, Tseng et al. 2016). These methods can use any information collected before drop out, and sources of unmeasured bias, to adjust results for a variety of patterns of missingness. The models for the survey will be based on the evolving CMO framework. We anticipate our Expert Groups will generate multiple, partially nested hypotheses, and variable selection will be conducted based on these. The relative plausibility of

these competing models will be evaluated using model-based fit criteria. In subsequent iterations, the Expert Groups will be aware of the best-fitting models, and will seek to evolve these. This evolution will occur largely at the level of model and sub-model, avoiding excessive reliance on the statistical significance of individual predictor variables, which may be influential in one context but not another.

***Qualitative Interviews***

We will interview key stakeholders online including approximately, 10-12 community members, 4-5 moderators, and 1-2 hosts and commissioners in each of our community cases (total sample up to 114). This will provide sufficient data to test our hypotheses, whilst also being manageable within the resources of the project.

Guided by realist interviewing methodology[87] [88], ***theoretical sampling for community members*** will be determined using an evolving sampling framework, and drawing on our survey and analysis of community posts, to identify which participants are needed to test our programme theories. This is likely to include at least one host, commissioner, and all moderators from our participating communities. We will sample across observers, super-users and regular community members. Where relevant data is available (e.g. SHaRON, Bipolar UK), our sample will include people who have been invited to join an online community but not done so (particularly targeting those from under-served groups), and those who have left the online community. All online communities work slightly differently. In many of our cases, people sign up individually, and it is not possible to know who was offered access to the community but chose not to engage. For SHaRON, hosted by Berkshire NHS Foundation Trust, only people referred by a clinician are offered access, so we can invite people referred but who chose not to engage to interview. Data are also collected on all members of Bipolar UK and all members are invited to join the e-community. We can access summary statistics of people who choose not to join the community, or who have withdrawn. Invitations can be sent internally to these members to invite them to take part. We will interview people who are invited to join online communities but did not, and people who withdraw to help us understand any feared anticipated outcomes, or experienced negative outcomes driving this.

***Interview topic guides*** will include an opportunity for participants to share what they understand the role of online communities to be, their impacts, and how they "work". The use of interviews allows us to contextualise communities more broadly and explore how participants came to be involved with the online community (or not), how their role within it evolved over time, their experiences to date, including why they left (if relevant), and if involvement in the community has led them to go on to access other forms of support. With our moderators and hosts / commissioners we will also explore their understanding of the role of moderators, how they make decisions regarding moderating behaviours, and how they are trained and supported in this role.

We will also include questions that directly interrogate our hypothesised programme theories[88]. Interview topic guides will evolve alongside theory and will be designed specifically to test our hypotheses including the role of moderators.

Interviews will be done iteratively to allow the theories and framework to develop during data collection. Interviews will be face-to-face or via secure video conferencing (Microsoft Teams), or telephone, recorded and transcribed.

***Analysis*** will first be done within case, and then integrated to compare between cases. Data will be managed in NVIVO. Consistent with the realist approach, analysis will be retroductive i.e. seek to identify the hidden causal forces underlying people's descriptions of their experiences in the forums. These will be coded into our hypothesised CMO configurations from Work-stream 1. They will add to, elaborate, refine, and refute CMOs as appropriate to develop our programme theory. Interviews will be coded and Interviews and initial coding will be done by our researchers, but the analysis will be developed through regular discussion with the co-app team and our Expert Groups.

***Analysis of community posts.***

The collection of posts from our online communities will result in very large linguistic datasets, amounting to millions of words in each case. To analyse these, we will use Corpus-based Discourse Analysis and Natural Language Processing, in which researchers at Lancaster University are world leaders[89, 90]. These methods make it possible to combine quantitative computer-aided analyses of large linguistic

datasets with in-depth qualitative analyses of selected texts or exchanges to specifically test our developing programme theories. These methods have been used successfully to improve understanding of many areas of health, including by our own team in dementia [91], public health [92], trigeminal neuralgia [93], cancer [94-96], end of life care [97, 98], psychosis [99], and methodologies [100].
Using two stages of computer-aided techniques selected from well-established techniques in corpus-based discourse analysis[101], natural language processing, and social network analysis, we will first describe and compare the size, shape, content, emotional tone and linguistic features of each of our community cases, and between subgroups of members such as moderators, super-users, and regular members. We can also look at how each of these features changes over time. In a second stage, we will then adapt methods to test our specific CMO configurations. Specific techniques we can draw on to do this are explained in Table 1.Stage 1 techniques help us describe the context, locate important or unusual patterns in the data, and direct us to specific posts or threads for further detailed analysis to specifically test our CMO configurations, typically using the Stage 2 techniques.

**Table 1- Overview of computational linguistic techniques to analyse community posts**

| **Stage 1 technique** | **Function** |
|---|---|
| Frequency and dispersion analyses | To obtain the frequencies and distribution of selected words, phrases or semantic domains in a corpus and its sub-sections. |
| Keyness analyses | to identify and study the words, phrases and semantic domains (i.e. concepts) characteristic of a particular (sub-)corpus, i.e. that are statistically significantly more frequent in one (sub-)corpus as opposed to others, e.g.: an online community vs. a general corpus of English; moderated vs. unmoderated communities; first posts vs. last posts; patients' posts vs. moderators' posts; etc. |
| Usage Fluctuation Analyses | a newly-developed tool at Lancaster University [94] to track changes in co-occurring words of critical interest over time, to capture continuities and discontinuities in meanings and associations |
| Summarisation or topic modelling | to generate short summaries of forums or parts of forums and locate the most important posts. We will use abstractive and extractive summarisation methods respectively along with topic modelling. These techniques will help to triangulate the results provided by quick gisting and visualisations provided by the corpus based keyness analysis. |
| Social network analysis | to extract the structure and shape of the network of people in the community, to see who is talking to whom, to classify user's participation style (observers, users, super-users) and how connected they are, their length of use of the forum, overall network density, and who are the influential posters. |
| User profiling | to estimate the relative usage and representativeness of people in each community by age and gender, country location and other demographic variables |
| **Stage 2 technique** | **Function** |
| Concordance analyses | to obtain all occurrences of selected words, phrases or instances of different semantic domains, and to understand how and why they are used, e.g.: words such as 'feel', 'daughter', and 'medication' or semantic domains such as 'Sad', 'Personal relationships' and 'Medicines and medical treatment'. We will extend this method to filter/sort concordance lines using metadata or linguistic annotation at individual post or thread level such as empathy, tone, size, or speed of replies |
| Collocation analyses | to study patterns of co-occurrence of words or phrases, which are well known to contribute to their meanings and associations, e.g. what words tend to co-occur with 'you', 'not' or 'worry' in different (sub-)corpora. The GraphColl tool in #LancsBox will be employed to study broader networks of collocates, e.g. the second-order collocates of particular words (or the collocates of collocates). |
| Selected qualitative analyses of individual texts or threads | The corpus tool ProtAnt makes it possible to rank texts in a (sub-)corpus in terms of their prototypicality, operationalised in terms of the frequency of keywords for the whole corpus. This will be used to objectively identify specific posts, threads or interactions typical of a particular (sub-)corpus, for detailed qualitative analysis (e.g. focusing on metaphors or personal narratives). |

| | |
|---|---|
| Sentiment analysis | To study change in emotion over time and across forums or threads, showing positive and negative emotions, and we will combine this with an analysis of risk indicator words and taboo words, and terms of emotional support or bullying. We will investigate the potential use of toxicity ratings such as Perspective.api although these will need augmenting with discourse level models to avoid 'out of context' errors [41], and to incorporate subjective reader/audience responses via measuring sentiment of replies and toxicity trajectories in threads. |

Analyses will be carried out using the latest versions of cutting-edge corpus tools developed at Lancaster University (#LancsBox, the Lancaster Stats Tools Online, lexiDB and Wmatrix), and, where appropriate, the corpus analysis tools available in AntLab, and Sketch Engine and state-of-the-art NLP toolkits such as NLTK, Spacy and Stanza.

Qualitative analyses will be employed to carry out in-depth studies of specific posts, interactions or threads selected to test our specific theories, and on the basis of the quantitative findings. This might include threads with particularly high or low frequencies of negatively valenced emotion words; posts or threads that were found to be prototypical of a whole community in linguistic terms; posts that received large numbers of replies; posts by individuals representing different levels of engagement or roles with the online community, or showing evidence of significant change in emotional state. The qualitative analyses can investigate, for example, what negatively valenced emotion-related words are used in a particular thread, how they are used (e.g. do they relate to the self, other people, situations, etc.), what interactional patterns they occur in (e.g. mutually supportive vs. conflictual), and whether and how users provide explanations for the relevant negative emotional states that are relevant to the research questions (e.g. the absence of a moderator, the anonymity of online forums, etc.).

Depending on the goal of each analysis, relevant concepts and frameworks will be drawn from qualitative discourse analysis, including, for example, the use of narratives to construct identities, the use of metaphors for the expression of intense emotions, the use of humour to construct in-groups and out-groups, the management of 'face' and relationships in advice seeking and advice giving, and the use of expressions of empathy, solidarity or conflict in interactions among users. These analyses will be directed at testing our CMOs by exploring users' understandings of mechanisms underlying the impact that participation in the online community has on themselves and others.

***Triangulation of data across methods and between cases***

Our novel interdisciplinary design has many benefits. All methods have limitations, and in combining our survey, interviews, and analysis of community posts, we can ensure that: we have a full range of perspectives (including those who choose not to take part in all aspects); we can purposively sample for interview on the basis of self-reported and real time evidence of specific impacts, or membership to specific subgroups; we can move between broad contextual data, such as the size and structure of the community, to detailed analysis of individual conversations. Regular analysis meetings with research team including our Expert Groups will ensure that we make the most of the opportunity for iterative analysis across our cases and between our researchers using different methods.

To illustrate how we can triangulate our research methods, to test and refine our explanatory theories, about the role of moderators (RQ2), and the impact of online communities on mental health outcomes (RQ1), we can consider the following specific (but theoretical) example.

RQ2: Based on our literature review and stakeholder interviews in phase 1, we might hypothesise that "the timing and style of a moderator interjection (**Outcome**) in an escalating toxic conversation will be a result of their understanding of what is happening and what the role of the moderator should be (**Mechanism** ) in this situation. This understanding will depend on the training/support/community rules they have received as a moderator (**Context**), which are themselves likely to be influenced by who is hosting the community e.g. NHS commissioned vs user-led charity (broader context)".

RQ1: We might then construct a CMO to understand how the timeliness and style of the moderator interjection then becomes the context influencing the mental health outcomes for the community members. So we might hypothesise that the timeliness and style of the moderator (**Context**) influences the extent to

which community members perceive themselves to be safe and cared for in the community (**Mechanism**), and that this perception determines their anxiety levels (**Outcome**).

By triangulating our research methods, we can test these CMOs in a number of ways. For example, we could use Sentiment Analysis to identify when difficult or toxic conversations are happening in the community. We could measure the timeliness of moderator activity and use (Im)politeness theory to describe the approach taken by the moderator in an interjection aimed at de-escalating tension. We can then interview the moderator to explore their reasoning behind managing these situations and talk through their thinking behind some of these specific examples (RQ1).

We can measure the effect of the moderator's interjection by using Keyness analyses combined with semantic tagging to compare the language used on the relevant thread prior to and following the interjection, including, for example, in terms of changes in the use of language related to anxiety levels. We will apply this approach to threads which differ in terms of the timing and characteristics of the moderator's intervention, and compare the size and nature of differences in language use prior to and following the intervention across the different threads. We can then use our survey methods to compare changes in levels of anxiety for individuals between communities and see if this is linked to average ratings of timeliness and sensitivity of moderator interventions for each community. We can use the survey to sample people who have shown extreme changes in anxiety (worsening and improving) and interview them to understand their lived experiences of the communities, exploring how safe they felt and what influenced this (RQ2).
Community data preparation and descriptive analysis will be completed alongside theory development in WS1. Methodological leads and RAs will attend bimonthly theory development meetings with our Expert Groups to understand how our programme theories are evolving. By month 18, the team will have a clear understanding of the programme theories to be tested in WS2. We will prioritise the order of testing specific theories based on perceived relevance, and logical order in the broader framework. Our methodological leads will test specific CMO configurations and share their findings at the bimonthly meetings with our Expert Groups. As in the realist review, we will look for evidence that supports, elaborates, and refutes our theories. All research methods have limitations, but used together we will have robust evidence across multiple cases, purposively sampled, and analysed from multiple perspectives using a range of mix-methods tools.

We anticipate presenting our final programme theories as a series of CMO configurations with corresponding narrative, including vignettes, to aid understanding of the main impacts of using online communities, and how these come about for, for who, and in what context. Following the recommendations that realist mechanisms should be viewed as comprising of both *changed resources* and *participant's reasoning/responses* to those resources, we will disaggregate the final programme theories into resources and reasoning. This disaggregation will enable the identification of the resource implications associated with the developed programme theories, which will be presented alongside the CMO configurations.

We will have separate CMOs configurations to understand key decision points in moderator behaviour to inform co-design of our e-learning resources and of the community of practice for moderators. We will also work closely with our Expert Groups and a visual artist to design a series of visual representations to maximise accessibility of our findings.

### *Rigour*

The validity of the findings is strengthened through triangulation of evidence across multiple sources of data, purposively sampled, and analysed from multiple perspectives using a range of mixed –methods approaches. Our methods will be systematic, clearly documented and reported following realist evaluation publication standards [102].
Our **Expert Groups** will be involved in the iterative process of interpreting findings from each of these methods to finalise our programme theories of how online mental health communities impact on mental health outcomes (**RQ1**) and the role of moderators (**RQ2)**.

**Work-stream 3- Co-designing and Implementing Theory Informed, Best-Practice Tools (2-34 months)**

In Work-stream 3, our theories will be used to underpin the co-design of best-practice tools and support for people involved in commissioning, hosting, moderating or using online mental health communities (**RQ3)**. Some outputs can be created before a finalised theory, (see outputs table) and so this work-stream also runs in parallel to WS1&WS2

***Co-design process***
This process will be led by PPI co-app Mat Rawsthorne who also has extensive published expertise in designing and facilitating co-design processes to develop mental health research questions [24], tools [71] and techniques [54]. Co-design is essential to ensuring that: the tools build on existing knowledge and practices; we draw on the most relevant stakeholder knowledge and expertise; users of the tools have a sense of ownership which will facilitate uptake and use; the tools are seen as credible by the wider community. The Co-Design group will comprise of approx. 8 (paid) volunteers from the PPI, Hosts and Impact Groups. This ensures that all members will be fully informed of the developing programme theory, and therefore this group can focus specifically on co-designing the tools to inform policy and practice.

The Co-design Group will be established at the very outset of the project (month 2) and will be co-facilitated by Rawsthorne and Lodge. The first discovery meeting will be face-to-face in which the members will be introduced and taken through two proven, evidence based methods which establish the focus and style of the group: 1. The Prosocial matrix[103] to agree ground rules amongst contributors by having them consider what's important to them and what they are willing to do to achieve that; 2. a TrustScapes [104] exercise to consider the worst possible outcomes from online communities and how to mitigate those risks to move closer to what matters to them. Subsequent workshops (min. of 10) will focus on developing and refining the theory informed tools and will be facilitated via an online whiteboard tool (Miro) which enables asynchronous contribution. The first tool to be established will be the Community of Practice for moderators, which will then serve as a sandpit in which to test subsequent tools in development.

***Candidate Tools- and Implementation***
All tools will be informed by the programme theory developed in Work-streams 1&2. They will be designed to accommodate regular updates as required in response to future refinements to the programme theory, advances in technology, and changes in the role online communities play in our lives. The exact nature of the tools will depend on the outcomes of the co-design process, but our pre-application work with our PPI, hosts and impact groups has suggested the candidate ideas listed below. All tools will be created as reusable learning objects built in Xerte that can be transported as SCORM objects to facilitate implementation (described below).

For people with mental health difficulties, referrers, and commissioners
- ***Knowledge exchange video/animation tools*** to widen access and promote greater uptake through better understanding of the role of online mental health communities and how they work. Aimed at commissioners, referrers, and the general population, these will describe how online communities work, possible benefits, for who, and how to identity safe and supportive communities.

For community moderators
- ***Community of Practice (CoP)*** for moderators which aims to build capacity through facilitating mutual engagement, joint enterprise, and shared practice[105]. Such support is crucial to continuing professional development, and to prevent compassion fatigue and burnout. CoPs cannot be "set-up" but come to life through the process of "thinking together"[106]. Our aim is to create the context that could facilitate this process, inviting all the moderators involved in our study, and providing a safe online space for the community to evolve from the outset of the study.
- ***E-learning curricula*** to train and support moderators in reflexive practice including: understanding the moderator role; ethics of moderation; encouraging activity; understanding mental health; spotting moments of change (introductory linguistic analysis); managing challenging situations; widening access and welcoming diversity; identifying and managing risks; looking after yourself; role of supervision & peer support; signposting; continuing professional development (CPD). The content will draw on

challenges described by moderators during our interviews, and practical case examples of ways to manage these. Suggested strategies will be piloted in a "sandpit" community by our co-design team, and Community of Practice moderators. Our training will be theory driven and focus specifically on issues relevant to moderation in health communities, distinguishing it from any existing moderator training we could identify that was aimed at online communities to build commercial brands (moderationgateway.com or discord.com/moderation).

For community providers/hosts, commissioners and policy makers

- ***Best practice design principles*** for online mental health communities to widen access, maximise positive impacts, and minimise harms. This guidance will provide 1) High level design principles to guide community design and moderation. 2) Practical guidance for implementing design features to improve usefulness and safety of online communities 3) Examples of best practice and case studies on community design and moderation practices.
- ***Resource requirements for implementation of best practice design principles*** based on the disaggregation of the programme theories into resources and reasoning. This will provide information on the resource configuration necessary to provide safe a and effective online community.
- ***Standardised community characterisation and evaluation framework*** based on the best practice design principles i.e. consistent and detailed information about each community which focuses on the features identified as crucial to the triggering of underlying causal mechanisms. Need for this framework has been identified by NHSE to support standardised evaluation and comparison of communities during commissioning.
- ***Methodological guidance on how to evaluate online mental health communities*** based on methodological insights from carrying out this study. This will provide guidance for both evaluation of effectiveness and considerations for future economic evaluations.

For further theory and practice development

- ***Programme theory*** and evidenced based tools for online mental health communities that have relevance to online health communities of other health conditions, and in other countries, and to offline peer to peer support.

**Implementation**

As the host site Berkshire NHS Foundation Trust will take a lead role in ensuring the best practice tools outputs will be easily discovered by our relevant audiences, openly accessible, free to use, and updated and maintained beyond the life of the project. They are a Digital Exemplar Trust with a vested interest in being identified as leading the field in online communities. They host SHaRON which has been set up across four other mental health trusts and adapted to support professionals working with children and young people in schools. SHaRON is promoted by NHSX and there is ongoing work to host additional SHaRONs across the UK and internationally.

The tools are targeted at moderators working in NHS, social care, voluntary sector, and commercial organisations. To ensure wide impact, Health Education England (represented by Henrietta Mbeah-Bankas, Head of Blended Learning and Digital Literacy) have agreed to promote the tools on their new online eLearning Hub with1.6 million users across health and social care (https://learninghub.nhs.uk). A national network of peer supporters working in charities, (represented by Duncan Marshall from Mind) will support design and implementation across charity sector organisations.

The best practice design principles, standardised community characterisation framework, and guidelines on evaluation need to be accessed by community hosts, commissioners, and by NHS England and NHSX to inform their digital support agenda. James Woollard (National Specialty Advisor for Digital Mental Health) and Jeremey Clark (Mental Health Policy Manager) will sit on our Impact Group. They will facilitate translation of our findings into a format that can inform the NICE Evidence Standards Framework for digital health technologies, and the support offered by the NICE META tool (Medtech Early Technical Assessment); and ensure this is consistent with the "safety by design framework" being developed on the basis of the UK Online Safety Bill.

The implementation of these standards is facilitated by organisations like ORCHA https://orchahealth.com. NHSX, identified ORCHA as a leading independent digital health evaluation and distribution organisation, working to promote safe digital adoption. ORCHA were instrumental in developing the NHS Apps library. Simon Leigh (R&D lead at ORCHA) will be part of our Impact Group and guide the translation of our programme theory into tools that can be used by ORCHA to support digital health providers to build better online health communities. ORCHA also train health professionals in understanding how to select health technologies, and our programme theory will inform this training. Finally, our Impact Group will also include Janet Ince, from our local mental health commissioning team, who will facilitate design and dissemination of our outputs to commissioning teams to improve decision making around commissioning of online communities.

The Oxford Academic Health Sciences Network have agreed to support a national launch of the co-designed tools, as part of a conference, webinar, and social media strategy. We have been adopted by our local North West Coast ARC, and linked into the national ARC Mental health Implementation Network to support an ongoing programme of implementation.
This implementation strategy will ensure that: mental health commissioners with have guidance on how to evaluate or compare existing online communities using a standardised framework; providers wanting to offer an online mental health community to their service offer will know what design features are important to ensure it safe and effective; community moderators will be able to access online training and peer support for this complex role; and people experiencing mental distress will have guidance on what kinds of online communities are likely to lead to what impacts for them, and how this might happen.

**Dissemination**

Our Expert Groups (PPI, Hosts, Impact) and Study Steering Group will also play a crucial role in guiding the dissemination of outputs. We will define our range of audiences, identify what each needs to know, and prepare targeted outputs appropriately created and written for this audience including Plain English summaries. Our protocols, and findings will be published in high impact journals (e.g. JMIR IF 5.03), and presented at academic and peer community conferences (e.g. Mindtech; NSUN). All Participants and relevant health, social, educational and academic communities will be kept updated via a study website, social media, and blogs from across the wider team.

**Equality diversity and inclusion**

Online communities have the potential to improve access to healthcare for groups traditionally under-served by services. Because they are not geographically bound, and do not require individuals to share their offline identity, they offer a route to support for people with rare or highly stigmatised difficulties, including mental health or physical/learning disabilities, and those living in remote areas, or close knit communities wanting anonymity.
However, we also know that online support can exclude those without access to the required technology (first level digital divide), skills to use it (second-level), or who do not derive the same benefits from use (third-level) exacerbating existing social inequalities in health [107, 108]. There is little data available to date to describe the populations using online communities, and how these differ from those accessing face to face support.
This study aims to address this “digital divide” in three ways. First by better describing demographics of who does and who does not use online forums, second, by developing a programme theory to understand how individual characteristics influence the experiences and impacts of online mental health communities, and third by ensuring this theory guides development of best practice tools that are inclusive by design. Informed by NIHR Guidance on how to improve inclusion of under-served groups in clinical research based on the INCLUDE project[109], we will describe the populations using our online community cases, and compare them to national data on population characteristics of those living with mental health difficulties, and those accessing service. In developing our programme theory, we will purposively sample people across demographic characteristics.

In Work-stream 3, our three Expert Groups will play a key role in the co-design of our theory informed best practice tools, and ensure direct path to impact. We will recruit experts to each of these groups who can

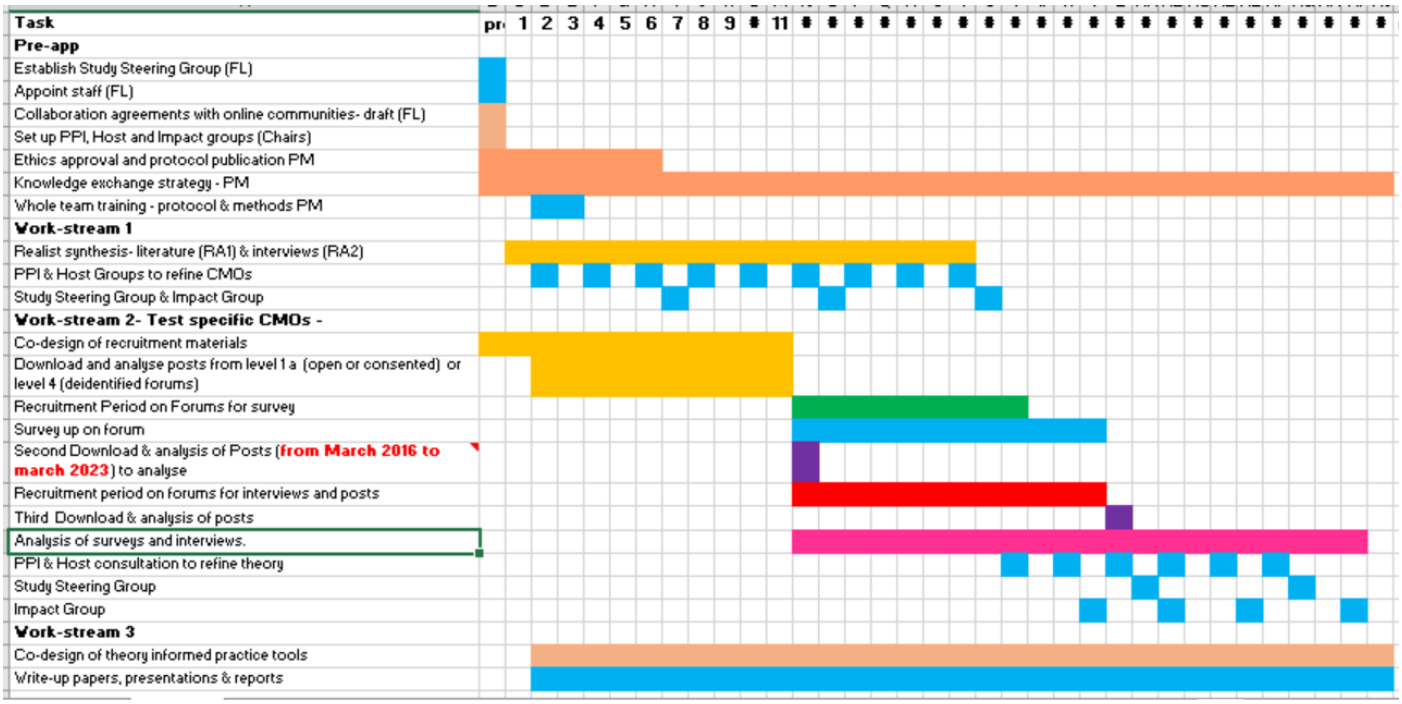
ensure that our tools maximise access, uptake and use by under-served populations. However, we acknowledge that peer support is inherently culturally determined and further work is needed to understand how online mental health communities might work in other cultures, such as low and middle income countries where they have huge potential to support limited healthcare provision especially across rural areas.

We will ensure our research is inclusive across age, gender, ethnicity, and disability and works to narrow the digital divide. In collaboration with the NWC ARC, we have used the Health Inequalities Assessment Tool https://www.hiat.org.uk/ and NIHR Guidance on how to improve inclusion of under-served groups (109) to further develop our strategy.

a) We will describe the samples of people who use online communities in our study cases by age, gender, and ethnicity, collected at sign up. Where possible we will compare this data to people who are invited, but chose not to use the communities, and to national population statistics. Data from those choosing not to engage is available through SHaRON and Bipolar UK.
b) We will purposively recruit people for our each of our Expert Groups, our Steering Committee, and our Co-design Group across a wide range of age, gender, ethnicity, disability, and digital engagement. We will use the same purposive sampling for participants to our stakeholder interviews in WS1 and theory testing interviewees in WS2. We will ensure all methods for involvement are accessible (including face to face meetings in community settings, and providing digital hardware where needed) in line with inclusive strategies in our previous research (24).
c) We will invite digitally inactive people to 3 in-person workshops to ensure their views inform the co-design process. We will widen our Co-design group in WS3 to 12 volunteers from across the Expert Groups.
d) We will design all outputs and our dissemination strategy with our Expert Groups so they are accessible to across age, gender, ethnicity, disability and the digital divide. We will write plain English summaries, translated into other languages (courtesy of https://onereach.ai/ ), and disseminate via non-digital platforms (radio, & print media) as well as digitally (websites, social media etc).
e) We anticipate challenges, particularly in engaging people not currently engaged in online communities or digitally active, and those from ethnic minority groups. To address this, we will work with organisations already engaging with these communities including The Good Things Foundation https://www.goodthingsfoundation.org/ , a social change charity for digital inclusion; and Sharing Voices https://sharingvoices.net/, a charity aiming to reduce mental health inequalities in Minority Ethnic communities.

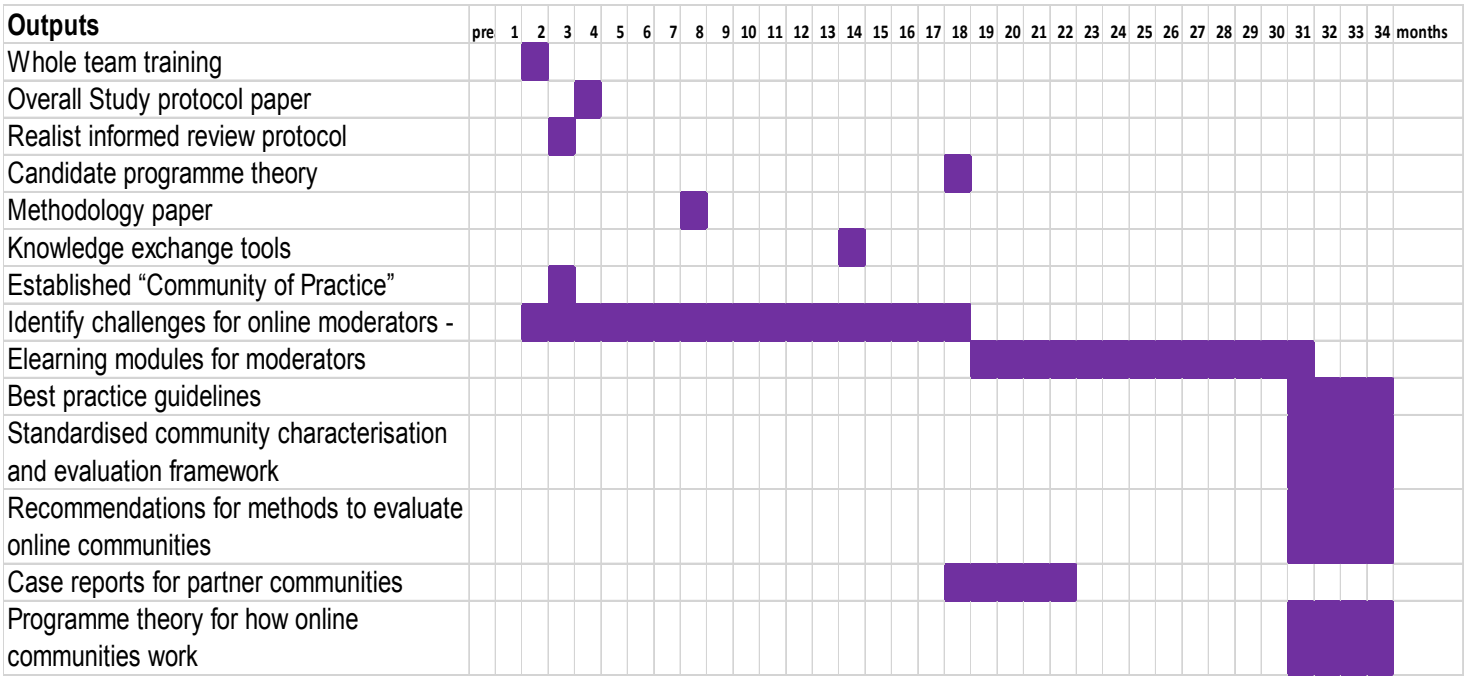
## Project / research timetable

All key tasks and timescales are shown in the Gantt chart below. These are estimates




The Work-streams run in parallel and co-design of all our outputs starts from month 2.

**Timing of Key Outputs are shown below.**




## Project Management

Chief Investigator (Lobban) will be responsible for delivering the project to time & budget and management & professional development of employed staff. The management team includes all co-apps, employed researchers (x2) and project manager. This group will meet weekly initially, moving to fortnightly as the study is established. Working sub groups led by Rayson (computational analysis of posts); Lobban (realist synthesis; impact group, qualitative interviews); Semino (linguistic analysis); Jones (survey & PPI /Co-design support); Rawsthorne and Lodge (PPI& co-design); Robinson (realist synthesis; ethics & governance; and hosts communication) will deliver specific tasks, and supervise of work by research and admin staff.

We have 3 Key Groups essential to our research design.
1. Our PPI Group (via Lodge & Rawsthorne) will meet every 2 months. They will be presented with data and in discussion with the research team, will iteratively develop and prioritise our programme theories.
2. Our Hosts Group (via Robinson) will include senior representatives from our partner communities. Also meeting every two months to input to the theory development and facilitate access to participant data.
3. Our Impact Group (via Lobban & Wise) will include senior policy makers, digital and mental health specialists, and commissioners. They will meet 6-monthly increasing to 3-monthly towards the end of the study to ensure the developing theory is informed by, and informs the future of peer online communities.
Our experience (supported by research [110]) tells us people feel more able to contribute in smaller homogenous groups, with less imbalances of power to contend. To ensure collaboration across groups, they are all facilitated by a member of management group (2 for PPI) who will be responsible for communication in both directions, and some joint meetings will be held as appropriate. Members from each of these 3 groups will be sampled to form the Co-design Group in Work-stream 3. An independent Study Steering group of methodological experts will meet at 6 monthly intervals, chaired by Prof Steven Gillard, an international expert in peer support interventions.

Due to the novel interdisciplinary design, and in line with good practice guidance to train the wider research team (including PPI group), we have costed to host a 2 day event at the start of the project to share study design, governance, and provide basic training in each of the research methods. This is crucial to building team relationships face to face, facilitating follow-up work to happen remotely (saving costs overall).

Microsoft Teams will be used for easy access to protocols, meeting agendas etc. Data will be stored on Lancaster University's approved IT systems and synthesised using NVivo.

**Ethics – Please refer to Data Management Plan.**

**Study end is defined as the last codesign workshop in which all best policy and practice tools are shared with the wider team. We anticipate this will take place in December 2024.**

**Intellectual Property**

Background IP is held by collaborating communities in the forum of a) community posts (copyright); b) design of community and training of moderators (protocols); c) expertise held by the stakeholders taking part in the Expert Groups (hosts, moderators and community members). These will be provided to Lancaster University to use for the purposes of this research only. Future IP will be the learning from the project. This will be owned by Berkshire NHS Trust as the research host, but we will recommend to NIHR that this should be exploited by all participating communities to maximise the impacts of the findings as widely as possible.

**Our Team** has expertise in (i) the lived experience of seeking mental health support to inform co-design research (Lodge, Rawsthorne) (ii) realist methodology and case study design (Rycroft-Malone, Lobban) (iii) designing, disseminating and evaluating online mental health support (Lobban, Jones, Wise), associated ethical and governance frameworks specific to digital mental health (Robinson, Lobban, Wise), and statistical analysis applied to complex social issues (Shryane); (iv) integrating quantitative and qualitative discourse analysis of online forum data (Semino); and (v) Natural Language Processing (Rayson). The CI (Lobban) has an established track record in delivering digital mental health studies to time and budget. Wise is senior project lead for SHaRON at Berkshire NHS Trust and is ideally placed to co-facilitate the Impact Group with Lobban. Jones and Lobban contribute to development of national policy through NICE and NHSE and delivery of services through HEE. The strengths of the team are in the high level of methodological expertise in all areas, the novel combination of disciplines, and the established working relationships between the internationally renowned Centre for Corpus Approaches to Social Science (CASS) and Spectrum Centre for mental health research, both at Lancaster and including joint supervised PhD students. We will appoint 2 postdoctoral researchers, one with skills in realist methodology and qualitative approaches, and the other with computational linguistic expertise. Our project manager will specialise in design, digital communication and online data management.

**Success criteria and barriers to proposed work**

Key milestones of success- and how we will mitigate risks:

1. Establish an effective inter-disciplinary team. We have spent considerable time pre-application, building relationships, and co-developing a structure that will ensure effective collaboration and efficient delivery. Expert Groups are all trained together at the outset, meet regularly throughout the project with clear tasks, and are facilitated by the core research ensuring clear lines of communication.
2. Develop a manageable range of initial programme theories in WS1 for testing in WS2. Rather than having tightly specified inclusion and exclusion criteria, realist synthesis methods actively encourage the researcher to draw on theories, and wider literature to develop initial programme theories. Our experience has taught us the importance of balancing breadth and depth in this process. Through regular consultation, we will ensure that sources of data are carefully selected for relevance, and that sufficient time is allocated to developing robust candidate theories.
3. Testing candidate theories in WS2 relies on access to community posts, and recruitment of participants to complete our surveys and take part in interviews. All our hosts have experience of being involved in research, are actively seeking the answers to our research questions, and have ready access to recruit sufficient participants for our study. We have drawn on our own extensive experience of running online health studies to design our recruitment, retention, ethical, and data management approaches. Shopping

vouchers, electronic data collection, and regular updates and dialogue via the online communities will enhance recruitment and retention. The Clinical Research Network have confirmed the online communities are eligible for NHS service support costs to support this activity (see Soecat). If any of our hosts decide not to take part for any reason, we have expressed interest from other online communities so we are confident we could recruit additional sites.
4. Finally, it is essential our research has a strong impact. We have involved key stakeholders at all levels who have a role in guiding the future design and delivery of online mental health communities, including NHSE, NHSX, HEE, Peer support charity networks, ORCHA, mental health commissioners. They are involved throughout the project to guide the design and dissemination of outputs.

1. McManus, S., et al., *Mental Health and Wellbeing in England: the Adult Psychiatric Morbidity Survey 2014*. 2016: NHS digital.
2. Ofcom, *Children and parents: Media use and attitudes report 2019*. 2020, Ofcom London, UK.
3. NHS Confederation, *Digital inclusion in mental health - a guide to help increase choice and improve access to digital mental health services*. 2020.
4. Gillard, S., *Peer support in mental health services: where is the research taking us, and do we want to go there?* 2019, Taylor & Francis.
5. Saha, K. and A. Sharma. *Causal factors of effective psychosocial outcomes in online mental health communities*. in *Proceedings of the International AAAI Conference on Web and Social Media*. 2020.
6. Duggan, M. and A. Smith, *6% of online adults are reddit users.* Pew Internet & American Life Project, 2013. **3**: p. 1-10.
7. Smith-Merry, J., et al., *Social connection and online engagement: insights from interviews with users of a mental health online forum.* JMIR mental health, 2019. **6**(3): p. e11084.
8. Spinzy, Y., et al., *Does the Internet offer social opportunities for individuals with schizophrenia? A cross-sectional pilot study.* Psychiatry research, 2012. **198**(2): p. 319-320.
9. Hanley, T., J. Prescott, and K.U. Gomez, *A systematic review exploring how young people use online forums for support around mental health issues.* Journal of Mental Health, 2019. **28**(5): p. 566-576.
10. Naslund, J., et al., *The future of mental health care: peer-to-peer support and social media, Epidemiology and Psychiatric Sciences, 25, 113-122*. 2016.
11. Coulson, N.S., E. Bullock, and K. Rodham, *Exploring the therapeutic affordances of self-harm online support communities: An online survey of members.* JMIR mental health, 2017. **4**(4): p. e44.
12. Brady, E., J. Segar, and C. Sanders, *"I always vet things": navigating privacy and the presentation of self on health discussion boards among individuals with long-term conditions.* Journal of medical Internet research, 2016. **18**(10): p. e274.
13. Griffiths, K.M., J. Reynolds, and S. Vassallo, *An online, moderated peer-to-peer support bulletin board for depression: user-perceived advantages and disadvantages.* JMIR Mental Health, 2015. **2**(2): p. e14.
14. Powell, J., N. McCarthy, and G. Eysenbach, *Cross-sectional survey of users of Internet depression communities.* BMC psychiatry, 2003. **3**(1): p. 19.
15. Prescott, J., A.L. Rathbone, and T. Hanley, *Online mental health communities, self-efficacy and transition to further support.* Mental Health Review Journal, 2020.
16. Bartlett, Y.K. and N.S. Coulson, *An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication.* Patient education and counseling, 2011. **83**(1): p. 113-119.
17. Bailey, E., et al., *Moderated online social therapy for young people with active suicidal ideation: qualitative study.* Journal of medical internet research, 2021. **23**(4): p. e24260.
18. Easton, K., et al., *Qualitative Exploration of the Potential for Adverse Events When Using an Online Peer Support Network for Mental Health: Cross-Sectional Survey.* JMIR mental health, 2017. **4**(4): p. e49.
19. Ziebland, S. and S. Wyke, *Health and illness in a connected world: how might sharing experiences on the internet affect people's health?* The Milbank Quarterly, 2012. **90**(2): p. 219-249.
20. Malik, S. and N.S. Coulson, *'They all supported me but I felt like I suddenly didn't belong anymore': an exploration of perceived disadvantages to online support seeking.* Journal of Psychosomatic Obstetrics & Gynecology, 2010. **31**(3): p. 140-149.

21. Zhang, X., et al., *Health information privacy concerns, antecedents, and information disclosure intention in online health communities.* Information & Management, 2018. **55**(4): p. 482-493.
22. Lobban, F., et al., *An online supported self-management toolkit for relatives of people with psychosis or bipolar experiences: the IMPART multiple case study.* Health Services and Delivery Research, 2020. **8**(37).
23. Commission., C.Q., *"2020 community mental health survey: statistical release."*. 2020.
24. Hollis, C., et al., *Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership.* The Lancet Psychiatry, 2018. **5**(10): p. 845-854.
25. Rethink Mental Illness. *80% of people living with mental illness say current crisis has made their mental health worse.* 2020; Available from: https://www.rethink.org/news-and-stories/news/2020/04/80-of-people-living-with-mental-illness-say-current-crisis-has-made-their-mental-health-worse/.
26. Sorkin, D.H., et al., *Rise in Use of Digital Mental Health Tools and Technologies in the United States During the COVID-19 Pandemic: Survey Study.* Journal of Medical Internet Research, 2021. **23**(4): p. e26994.
27. NHS., *The NHS Longterm Plan.* 2019.
28. England., H.E., *The Competence Framework for Mental Health Peer Support Workers.* 2021.
29. King, A.J. and M.B. Simmons, *A systematic review of the attributes and outcomes of peer work and guidelines for reporting studies of peer interventions.* Psychiatric services, 2018. **69**(9): p. 961-977.
30. UK., G., *The UK ONline Safety Bill* 2021.
31. NHSX., *Digital Technology Assessment Critiera.* 2021.
32. NICE., *Evidence standards framework for digital health technologies.* 2021.
33. Griffiths, K.M., et al., *The effectiveness of an online support group for members of the community with depression: a randomised controlled trial.* PloS one, 2012. **7**(12): p. e53244.
34. Marchant, A., et al., *A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown.* PloS one, 2017. **12**(8): p. e0181722.
35. Dunlop, S.M., E. More, and D. Romer, *Where do youth learn about suicides on the Internet, and what influence does this have on suicidal ideation?* Journal of child psychology and psychiatry, 2011. **52**(10): p. 1073-1080.
36. Moore, D. and S. Ayers, *Virtual voices: social support and stigma in postnatal mental illness Internet forums.* Psychology, health & medicine, 2017. **22**(5): p. 546-551.
37. Rodgers, R.F., et al., *A meta-analysis examining the influence of pro-eating disorder websites on body image and eating pathology.* European Eating Disorders Review, 2016. **24**(1): p. 3-8.
38. Krasodomski-Jones, A., *Everything in moderation: platforms, communities and users in a healthy online environment.*
39. Sibilla, F. and T. Mancini, *I am (not) my avatar: A review of the user-avatar relationships in Massively Multiplayer Online Worlds.* Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 2018. **12**(3).
40. Suler, J., *The online disinhibition effect.* Cyberpsychology & behavior, 2004. **7**(3): p. 321-326.
41. SMITH, J., et al., *A picture of health march 2021: Measuring the comparative health of online spaces*, DEMOS, Editor. 2021.
42. Smith, J., E. Jones, and E. Judson, *What's in a name? A forward view of anonymity online.*
43. Morriss, R., et al., *A direct to public peer support programme (Big White Wall) versus web-based information to aid self-management of depression and anxiety (The REBOOT study): results and challenges of an automated randomised controlled trial.* Journal of Medical Internet Research.
44. Park, A. and M. Conway, *Longitudinal changes in psychological states in online health community members: understanding the long-term effects of participating in an online depression community.* Journal of medical Internet research, 2017. **19**(3): p. e71.
45. Pruksachatkun, Y., S.R. Pendse, and A. Sharma. *Moments of change: Analyzing peer-based cognitive support in online mental health forums*. in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019.
46. Wadden, D., et al., *The Effect of Moderation on Online Mental Health Conversations.* arXiv preprint arXiv:2005.09225, 2020.
47. Grimmelmann, J., *The virtues of moderation.* Yale JL & Tech., 2015. **17**: p. 42.
48. Atanasova, S., T. Kamin, and G. Petrič, *Exploring the benefits and challenges of health professionals' participation in online health communities: Emergence of (dis) empowerment processes and outcomes.* International journal of medical informatics, 2017. **98**: p. 13-21.
49. Watson, E., *The mechanisms underpinning peer support: a literature review.* Journal of Mental Health, 2017.

50. Gillard, S., et al., *Developing a change model for peer worker interventions in mental health services: a qualitative research study.* Epidemiology and Psychiatric Sciences, 2015. **24**(5): p. 435.
51. Wang, D., J. Weeds, and I. Comley. *Improving Mental Health using Machine Learning to Assist Humans in the Moderation of Forum Posts*. in *HEALTHINF*. 2020.
52. Sharma, A., et al., *Towards Facilitating Empathic Conversations in Online Mental Health Support: A Reinforcement Learning Approach.* arXiv preprint arXiv:2101.07714, 2021.
53. Ruckenstein, M. and L.L.M. Turunen, *Re-humanizing the platform: Content moderators and the logic of care.* new media & society, 2020. **22**(6): p. 1026-1042.
54. Rawsthorne, M., et al. *ExTRA: Explainable Therapy-Related Annotations*. in *2nd Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence*. 2020.
55. Joglekar, S., et al., *How online communities of people with long-term conditions function and evolve: network analysis of the structure and dynamics of the asthma UK and British lung foundation online communities.* Journal of medical Internet research, 2018. **20**(7): p. e238.
56. Jagfeld, G., et al., *Understanding who uses Reddit: Profiling individuals with a self-reported bipolar disorder diagnosis.* arXiv preprint arXiv:2104.11612, 2021.
57. Chew, R., et al., *Predicting Age Groups of Reddit Users Based on Posting Behavior and Metadata: Classification Model Development and Validation.* JMIR Public Health and Surveillance, 2021. **7**(3): p. e25807.
58. Yin, R.K., *Case study research and applications: Design and methods*. 2017: Sage publications.
59. Pawson, R. and N. Tilley, *Realistic evaluation*. 1997: sage.
60. Emmel, N., et al., *Doing realist research*. 2018: Sage.
61. Leamy, M., et al., *Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis.* The British Journal of Psychiatry, 2011. **199**(6): p. 445-452.
62. Champion, V.L. and C.S. Skinner, *The health belief model.* Health behavior and health education: Theory, research, and practice, 2008. **4**: p. 45-65.
63. Davidson, L., et al., *Peer support among adults with serious mental illness: a report from the field.* Schizophrenia bulletin, 2006. **32**(3): p. 443-450.
64. Machin, K., *8 Peer Support Training.* Peer Support in Mental Health, 2019: p. 99.
65. Bandura, A. and D.C. McClelland, *Social learning theory*. Vol. 1. 1977: Englewood cliffs Prentice Hall.
66. McMillan, D.W. and D.M. Chavis, *Sense of community: A definition and theory.* Journal of community psychology, 1986. **14**(1): p. 6-23.
67. Hatfield, E., J.T. Cacioppo, and R.L. Rapson, *Emotional contagion.* Current directions in psychological science, 1993. **2**(3): p. 96-100.
68. Shearn, K., et al., *Building Realist Program Theory for Large Complex and Messy Interventions.* International Journal of Qualitative Methods, 2017. **16**(1): p. 1609406917741796.
69. Flynn, R., et al., *Developing an initial program theory to explain how patient-reported outcomes are used in health care settings: methodological process and lessons learned.* International Journal of Qualitative Methods, 2020. **19**: p. 1609406920916299.
70. Morriss, R., et al., *Outcomes of a public health campaign and automated randomised controlled trial of a direct to public peer support programme (Big White Wall) versus web-based information to aid self-management of depression and anxiety (The REBOOT study).*
71. Craven, M.P., et al., *Try to see it my way: exploring the co-design of visual presentations of wellbeing through a workshop process.* Perspectives in public health, 2019. **139**(3): p. 153-161.
72. Geller, B., et al., *One-year recovery and relapse rates of children with a prepubertal and early adolescent bipolar disorder phenotype.* Am J Psychiatry, 2001. **158**(2): p. 303-5.
73. NSUN. *4pi National Involvement Standards*. 14/12/2020]; Available from: https://www.nsun.org.uk/4pi-involvement-standards.
74. Involve, N. *Guidance on co-producing a resarch project* 2018 14/12/2020]; Available from: https://www.invo.org.uk/wp-content/uploads/2019/04/Copro_Guidance_Feb19.pdf.
75. Pawson, R., et al., *Realist synthesis: an introduction.* Manchester: ESRC Research Methods Programme, University of Manchester, 2004.
76. Rycroft-Malone, J., et al., *Realist synthesis: illustrating the method for implementation research.* Implementation Science, 2012. **7**(1): p. 33.

77. Jagosh, J., *Realist synthesis for public health: building an ontologically deep understanding of how programs work, for whom, and in which contexts.* Annual review of public health, 2019.
78. Bonell, C., et al., *'Dark logic': theorising the harmful consequences of public health interventions.* J Epidemiol Community Health, 2015. **69**(1): p. 95-98.
79. Wong, G., et al., *RAMESES publication standards: realist syntheses.* BMC medicine, 2013. **11**(1): p. 21.
80. Kaylor-Hughes, C.J., et al., *Direct to public peer support and e-therapy program versus information to aid self-management of depression and anxiety: protocol for a randomized controlled trial.* JMIR research protocols, 2017. **6**(12): p. e231.
81. Powell, J., et al., *Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial.* Journal of medical Internet research, 2013. **15**(1): p. e2.
82. Spitzer, R.L., et al., *A brief measure for assessing generalized anxiety disorder: the GAD-7.* Archives of internal medicine, 2006. **166**(10): p. 1092-1097.
83. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure.* Journal of general internal medicine, 2001. **16**(9): p. 606-613.
84. Robinson, J., et al., *Why are there discrepancies between depressed patients' Global Rating of Change and scores on the Patient Health Questionnaire depression module? A qualitative study of primary care in England.* BMJ open, 2017. **7**(4).
85. Maxwell, S.E. and D.A. Cole, *Bias in cross-sectional analyses of longitudinal mediation.* Psychological methods, 2007. **12**(1): p. 23.
86. Rabe-Hesketh, S., A. Skrondal, and A. Pickles, *Generalized multilevel structural equation modeling.* Psychometrika, 2004. **69**(2): p. 167-190.
87. Pawson, R., *Theorizing the interview.* British Journal of Sociology, 1996: p. 295-314.
88. Manzano, A., *The craft of interviewing in realist evaluation.* Evaluation, 2016. **22**(3): p. 342-360.
89. Baker, P., et al., *A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press.* Discourse & society, 2008. **19**(3): p. 273-306.
90. Lancaster University, *The Centre for Corpus Approaches to Social Science is an ESRC-funded research centre (grant references: ES/K002155/1, ES/R008906/1) located at Lancaster University and operating in partnership with the University Centre for Computer Corpus Research on Language (UCREL) and the Academy of Social Sciences*.
91. Sutcliffe, A., et al., *Known and unknown requirements in healthcare.* Requirements engineering, 2020. **25**(1): p. 1-20.
92. Alsudias, L. and P. Rayson. *COVID-19 and Arabic Twitter: How can Arab World Governments and Public Health Organizations Learn from Social Media?* in *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*. 2020.
93. Semino, E., *Metaphorical descriptions of pain on a trigeminal neuralgia forum: pushing the boundaries of cognitive linguistics.* 2019.
94. Semino, E., et al., *The online use of Violence and Journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study.* BMJ supportive & palliative care, 2017. **7**(1): p. 60-66.
95. Semino, E., et al., *Metaphor, cancer and the end of life: A corpus-based study*. 2017: Routledge.
96. Hendricks, R.K., et al., *Emotional implications of metaphor: Consequences of metaphor framing for mindset about cancer.* Metaphor and Symbol, 2018. **33**(4): p. 267-279.
97. Potts, A. and E. Semino, *Healthcare professionals' online use of violence metaphors for care at the end of life in the US: a corpus-based comparison with the UK.* Corpora, 2017. **12**(1): p. 55-84.
98. Demmen, J., et al., *A computer-assisted study of the use of violence metaphors for cancer and end of life by patients, family carers and health professionals.* International Journal of Corpus Linguistics, 2015. **20**(2): p. 205-231.
99. Collins, L.C., et al., *Corpus Linguistics and Clinical Psychology: Investigating 'personification' in first-person accounts of voice-hearing.* International Journal of Corpus Linguistics. , accepted.
100. Semino, E., *Corpus linguistics and metaphor.* The Cambridge handbook of cognitive linguistics, 2017: p. 463-476.
101. McEnery, T. and A. Hardie, *Corpus linguistics: Method, theory and practice*. 2011: Cambridge University Press.
102. Wong, G., et al., *RAMESES II reporting standards for realist evaluations.* BMC medicine, 2016. **14**(1): p. 1-18.
103. Atkins, P.W., D.S. Wilson, and S.C. Hayes, *Prosocial: using evolutionary science to build productive, equitable, and collaborative groups*. 2019: New Harbinger Publications.

104. Ito-Jaeger, S., M. Rawsthorne, and E. Perez-Vallejos, *TrustScapes: Participatory tools for canvassing stakeholders' concerns and ideas on potential applications of artificial intelligence* Journal of Responsible Technology, 2021 (in press)
105. Wenger, E., *Communities of practice: Learning, meaning, and identity*. 1999: Cambridge university press.
106. Pyrko, I., V. Dörfler, and C. Eden, *Thinking together: What makes Communities of Practice work?* Human Relations, 2017. **70**(4): p. 389-409.
107. Wei, K.-K., et al., *Conceptualizing and testing a social cognitive model of the digital divide.* Information Systems Research, 2011. **22**(1): p. 170-187.
108. Van Deursen, A.J. and J.A. Van Dijk, *The digital divide shifts to differences in usage.* New media & society, 2014. **16**(3): p. 507-526.
109. Research., N.I.f.h., *Improving inclusion of under-served groups in clinical research: Guidance from the NIHR INCLUDE project.* 2020.
110. Racine, E., et al., *'It just wasn't going to be heard': A mixed methods study to compare different ways of involving people with diabetes and health-care professionals in health intervention research.* Health Expectations, 2020. **23**(4): p. 870-883.
111. Franzke, A.S., et al., *Internet research: ethical guidelines 3.0.* 2020.
112. Society, B.P., *Ethics guidelines for internet-mediated research.* Leicester, UK: British Psychological Society, 2017.
113. Seale, C., et al., *Interviews and internet forums: a comparison of two sources of qualitative data.* Qualitative health research, 2010. **20**(5): p. 595-606.
114. Jowett, A., *A case for using online discussion forums in critical psychological research.* Qualitative Research in Psychology, 2015. **12**(3): p. 287-297.
115. Roberts, L.D., *Ethical issues in conducting qualitative research in online communities.* Qualitative Research in Psychology, 2015. **12**(3): p. 314-325.
116. Chandrasekharan, E., et al., *Quarantined! Examining the Effects of a Community-Wide Moderation Intervention on Reddit.* arXiv preprint arXiv:2009.11483, 2020.
117. Chandrasekharan, E., et al., *You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech.* Proceedings of the ACM on Human-Computer Interaction, 2017. **1**(CSCW): p. 1-22.
118. Thomas, P.B., et al., *Behavior Change in Response to Subreddit Bans and External Events.* arXiv preprint arXiv:2101.01793, 2021.

Extra refs from Nick

119: Robinson H, Appelbe D, Dodd S, Flowers S, Johnson S, Jones SH, Mateus C, Mezes B, Murray E, Rainford N, Rosala-Hallas A, Walker A, Williamson P, Lobban F. (2020). Methodological Challenges in Web-Based Trials: Update and Insights From the Relatives Education and Coping Toolkit Trial, JMIR Ment Health 2020;7(7):e15878doi: 10.2196/15878 PMID: 32497018 PMCID: 7395253

120: Villanti A.C., et al. (2020). Recruiting and Retaining Youth and Young Adults in the Policy and Communication Evaluation (PACE) Vermont Study: Randomized Controlled Trial of Participant Compensation. J Med Internet Res, 22(7):e18446. DOI: 10.2196/18446