*Corresponding Author:

Ruth Gwernan-Jones, University of Exeter Medical School, Heavitree Road, Exeter EX2 1LU

A worked example of initial theory-building: PARTNERS2 collaborative care for people who have experienced psychosis in England

Ruth Gwernan-Jones[a*], Nicky Britten[a], Jon Allard[b], Elina Baker[a], Laura Gill[b], Helen Lloyd[b], Tim Rawcliffe[c], Ruth Sayers[d], Humera Plappert[e], John Gibson[e], Michael Clark[f], Maximillian Birchwood[g], Vanessa Pinfold[d], Siobhan Reilly[h], Linda Gask[i], Richard Byng[b]

[a]Institute of Health Research, University of Exeter Medical School, UK

[b]University of Plymouth, Faculty of Medicine and Dentistry, UK

[c]Lancaster Care NHS Foundation Trust, UK

[d]The McPin Foundation, UK

[e]University of Birmingham Primary Care Clinical Sciences, UK

[f]Personal Social Services Research Unit, London School of Economics, UK

[g]Warwick Medical School, UK

[h]Division of Health Research, Faculty of Health and Medicine, Lancaster University, UK

[i]Faculty of Life Sciences, University of Manchester, UK

Abstract

In this paper we present an exemplar of the initial theory-building phase of Theory-driven Evaluation (TDE) for the PARTNERS2 project, a collaborative care intervention for people with experience of psychosis in England. Initial theory-building involved analysis of literature, interviews with key leaders, and focus groups with service users. The initial programme theory (IPT) was developed from these sources in an iterative process between researchers and stakeholders (service users, practitioners, commissioners) involving four activities: articulation of 442 explanatory statements systematically developed using realist methods; debate and consensus; communication; interrogation.

We refute two criticisms of TDE of complex interventions. We demonstrate how the process of initial theory-building made a meaningful contribution to our complex intervention in five ways. Although time consuming, it allowed us to develop an internally coherent and well documented intervention.

This study and the lessons learnt provide a detailed resource for other researchers wishing to build theory for TDE.

Keywords: Programme theory development; theory-driven evaluation; complex interventions; collaborative care; personal recovery; psychosis

Abstrait

Dans cet article, nous présentons un exemple de la phase initiale d'élaboration de la théorie de l'évaluation basée sur la théorie (TDE) pour le projet PARTNERS2, une intervention de soins en collaboration destinée aux personnes ayant une expérience de la psychose en Angleterre. L'élaboration initiale de la théorie a impliqué une analyse de la littérature, des entretiens avec des dirigeants clés et des groupes de discussion avec des utilisateurs de services. La théorie initiale du programme (TPI) a été élaborée à partir de ces sources dans le cadre d'un processus itératif entre chercheurs et parties prenantes (utilisateurs de services, praticiens, commissaires) comprenant quatre activités: articulation de 442 déclarations explicatives systématiquement développées à l'aide de méthodes réalistes; débat et consensus; la communication; interrogatoire.

Nous réfutons deux critiques du TDE d'interventions complexes. Nous montrons comment le processus de construction initiale de la théorie a contribué de manière significative à notre intervention complexe de cinq manières. Bien que prenant beaucoup de temps, cela nous a permis de développer une intervention interne cohérente et bien documentée.

Cette étude et les leçons apprises fournissent une ressource détaillée aux autres chercheurs souhaitant élaborer une théorie pour le TDE.

Keywords: Développement de la théorie du programme; évaluation théorique; interventions complexes; soins en collaboration; récupération personnelle; psychose

Introduction

Evaluations of complex interventions have expanded from a focus on what works to attempting to understand the complexities of how an intervention does or does not work. Theory-driven evaluation (TDE) is one approach to answering complex questions about complex interventions, and in this paper we focus on the initial theory-building phase of TDE for the PARTNERS2 project, a collaborative care intervention for people with experience of psychosis in England.

TDE approaches initiated in the 1950s (Kirkpatrick 1959) and 1980s (Chen 1990) were taken up in the mid-1990s in an attempt to make sense of substantial government investment in interventions to improve local communities that resulted in disappointing, and difficult to understand, outcomes (Connell et al., 1995). More recently, the Medical Research Council's (MRC) framework for complex interventions (2008) has been criticized for failing to include theory-driven approaches in its guidelines (Anderson, 2008). This has been at least partially addressed by MRC guidance for process evaluation (Moore et al., 2015).

Along with a stronger mandate for theorising complex interventions, what is meant by evaluators' use of "theory" has become increasingly complex and confusing. Different kinds of theory in TDE, how they differ from each other, and how they are best used, is not always clear, although a few authors have attempted to clarify such issues. Partly to blame is the use of similar vocabulary for different things, and different vocabulary for similar things (Stame 2004; Blamey and MacKenzie 2007; Leeuw and Donaldson 2015). In the following paragraphs, we will clarify our understanding of, chosen vocabulary for, and aspects of TDE we will be describing in this paper.

Early in the history of TDE Suchman (1967) distinguished between programme theory and implementation theory: programme theory describes and explains how intervention activities bring about the desired effects of the intervention; whereas implementation theory explains how to put the intervention activities successfully into action. In Blamey and MacKenzie's (2007) exploration of Realistic Evaluation (Pawson and Tilley 1997) and its relationship to a Theories of Change approach (Connell et al., 1995), Blamey and MacKenzie associate Theories of Change with more of an emphasis on implementation theory, and Realistic Evaluation with a focus on programme theory. Blamey and MacKenzie do recognise that the two approaches overlap, since both aim to theorise how an intervention will work and to use theory to inform how the intervention should be evaluated, and both approaches acknowledge the importance of implementation *and* programme theory. However, the authors conclude that the two approaches elicit different kinds of information from stakeholders, are best used to articulate different aspects of an intervention, and generate different kinds of theory. We propose in addition that whether related to practice or implementation, realist approaches aim to understand how 'invisible' (Lacouture et al., 2015) and often universal mechanisms operate whereas Theories of Change is more situational. In this paper, we describe a theory-building process which draws from both realist and Theories of Change approaches, to develop theories of implementation alongside those that predict how the intervention will create its effects.

A different clarification of the use of theory in TDE is found in a review of the use of theory in interventions reported in the journal *Evaluation*. Leeuw and Donaldson (2015) clarify two

typologies: Typology 1 consists of theories of policy makers, stakeholders and evaluators underlying their professional work in making policies and doing evaluations; Typology 2 consists of scientific theories capable of contextualizing and explaining the consequences of policies, programmes and evaluators' actions. The typologies apply to both programme and implementation theory.

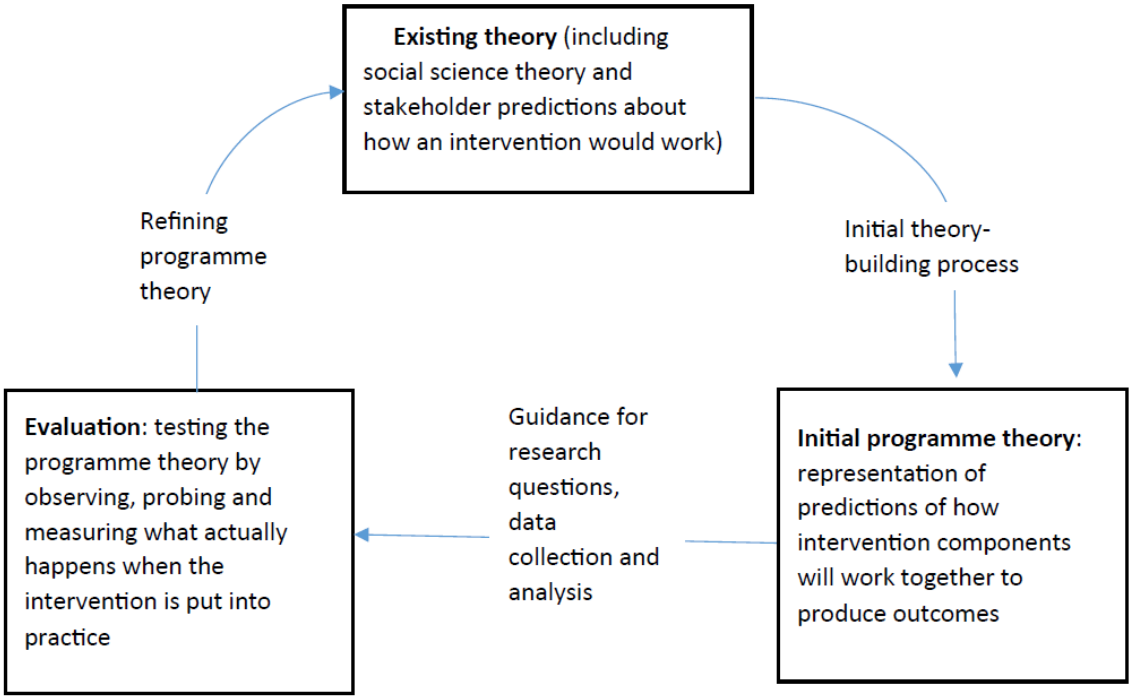
Typology 1 represents the rationales and expectations of stakeholders and researchers. Because stakeholders are already embedded in specific contexts, they are likely to have valuable, contextualised knowledge to which researchers are not privy. Developing implementation and programme theory from stakeholder knowledge is therefore considered an important means of linking theory and context (Moore and Evans, 2017). Leeuw and Donaldson (2015) detail numerous possibilities for TDE within Typology 1 to include, in addition to programme and implementation theory: theories of change, theories in use, logic models, logical frameworks, theory/anti-theory and evaluation theories. These all have in common their origin in the perceptions about how the intervention works of the stakeholders, researchers and evaluators involved.

By contrast, Typology 2 represents existing, more abstract social science theories. These provide theoretical explanation at the social and institutional level. Medical Research Council guidance suggests drawing on existing evidence and theory, and supplementing this with primary data (Craig et al., 2006). This corresponds to findings by Leeuw & Donaldson (2015) where almost half of the studies they reviewed synthesised both stakeholder/researcher theory and existing scientific theory to create a 'plausible' intervention theory. The authors suggest that combining the two typologies represents the most robust approach, and this is the approach we take in our initial theory-building phase.

Because both programme and implementation theory are relevant to most interventions, Weiss (1997) suggests a term ('theories of change evaluation') that refers to both. More recently, authors adopt the term 'program theory' in this more general sense to represent the specification of how various intervention components interact with each other to produce intervention effects, which includes implementation theory (Coryn 2011; Funnell and Rogers 2011), and this will be the sense with which we use 'programme theory' in the remainder of the paper.

Finally, we wish to draw attention to different stages of theory development within the cycle of TDE described above (see Figure 1) which can be easily confused because of the importance of programme theory throughout. The initial theory-building process involves drawing from existing theory, including the implicit ideas of stakeholders and existing social science theory, and synthesizing those relevant to the intervention to create an initial programme theory (IPT). This IPT is then used to guide the design of the intervention evaluation. Findings from the evaluation are used to refine the programme theory,

Figure 1. The theory-driven evaluation cycle. This paper discusses processes retrospectively identified between the *Existing theory* and *Initial programme theory* stages of TDE during the initial theory-building process of the PARTNERS2 intervention.



contributing further to existing theory. This paper will focus on processes retrospectively identified that relate to the *initial theory-building* process during the PARTNERS2 project.

Not all evaluation scientists are convinced that programme theory is beneficial. Two objections summarised by Coryn et al., (2011) are that: 1) explication of programme theory is unnecessary because it is often not used in any meaningful way, and 2) since developing high quality programme theories is often not feasible, and poor quality programme theories can be counter-productive, conducting TDE is a waste of valuable resources. Although there is a growing body of work in the field of TDE that debates what should be done to develop a useful programme theory, there are few examples that explore the role of this approach within specific projects. Coryn et al., (2011) conclude their review of projects employing TDE by calling for "exemplars, including reports of successes and failures, methods and analytic techniques" (p216) after finding a paucity of evidence either to support or contradict claims made by both critics of or advocates for TDE. Inquiry into the method of using theory during evaluation of interventions is rare (Brand et al., 2018), however scrutiny of the theory-building process is perhaps even more so. We were only able to find two detailed accounts (Pearson et al., 2015; Shearn et al 2017), both of which discussed theory-building specifically in relation to realist approaches. In this paper, although we draw from realist approaches as one aspect of initial theory-building (see section below 'Articulation drawing from realist approaches'), we present an exemplar of initial theory-building that draws more widely across both Theories of Change and realist approaches, because we develop theories of implementation alongside those that predict how the intervention will create its effects.

The PARTNERS2 project is funded by a UK National Institute of Health Research 5-year Programme Grant to include stages of theory-building (initial and refining), formative evaluation, and RCT with process evaluation informed by the programme theory developed

in the earlier stages of the research. The PARTNERS project involved a phase of initial theory-building in order to develop an intervention that would later be evaluated at pilot and trial stages. Although we test our IPT during later stages of the project, and this is reported elsewhere (Baker et al., 2019), this paper focuses on the initial theory-building stage only. As we noted above, there is a paucity of published articles articulating and reflecting on this crucial stage of TDE.

In creating an IPT we aimed to define and develop the key components of collaborative care for people with a diagnosis of schizophrenia or bipolar in an English primary care context. In doing this we aimed to conserve fundamental principles of collaborative care, including components and elements for which there was evidence of likely benefit, and adapting other components to make them optimal for people with experience of psychosis.

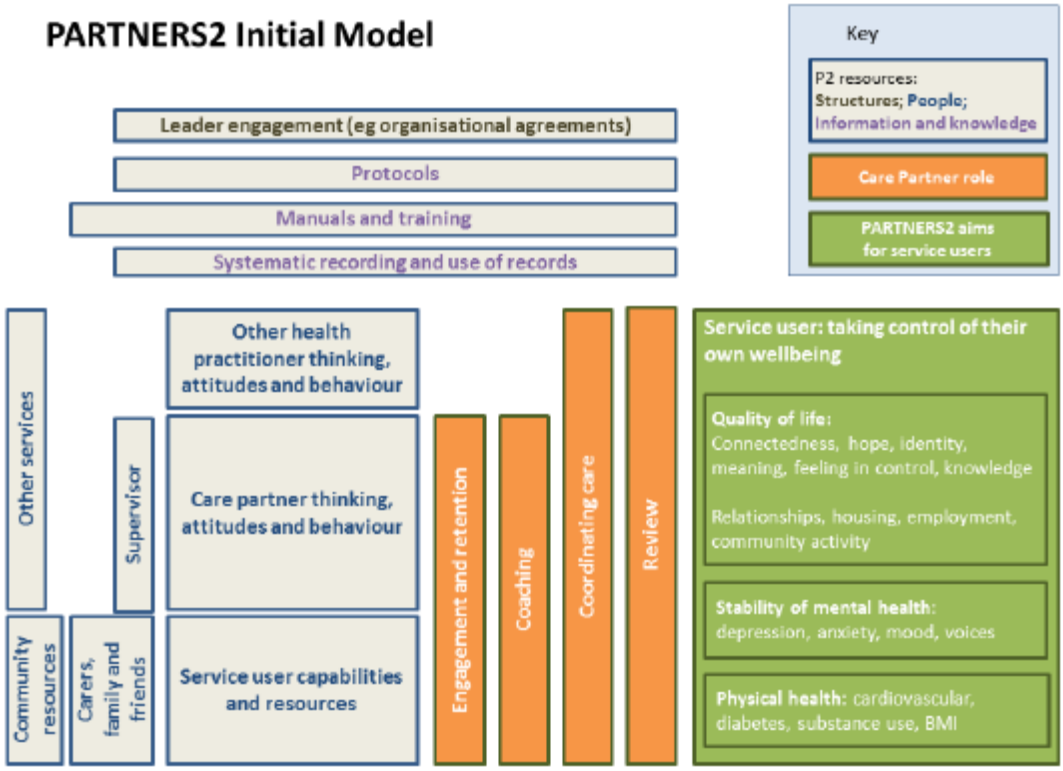
We adopted a theory-driven approach to support evaluation of a number of elements of complexity in the intervention, including multiple components, two targeted levels for change, multiple outcomes, tailoring of the intervention to individual recipients, and intervention sites involving multiple institutional systems, so complexity of context. In our initial theory-building processes, we explored both how the intervention should be implemented, and how it would bring about its outcomes. Our aim in this paper is to provide a worked example and to contribute to the debate about the value of theory in evaluation, by describing the initial theory-building activities with which we engaged, and by reflecting on how these activities impacted the IPT creation and content.

Developing the PARTNERS2 Initial Programme Theory

We developed the IPT in an iterative process from April 2014 to October 2015. The PARTNERS2 intervention aimed to improve physical health and wellbeing, and to stabilize mental health, of people living with psychosis in England. The IPT described how the intervention would achieve this, through collaborative care approaches such as a multi-professional approach to care (an experienced mental health professional called a ‘care partner’ sited within primary care), appropriate psycho-social intervention (sign-posting/referral and a coaching approach to individualized care), and regular/systematic monitoring and improved inter-professional communication (see Figure 2 for a graphic representation of the IPT). We framed these approaches using the concept widely used in mental health care in England of ‘personal recovery’, and coaching principles in order to support service users to be more active in managing their mental and physical health, and to orientate care around service user priorities.

To set the programme theory in context, we describe here its progress over the time of the PARTNERS2 grant period, although we only describe in detail processes of initial theory-building in this paper. The IPT was operationalised in our pilot intervention that was formatively evaluated from November 2016 to April 2017. The formative evaluation contributed to further refinement of the IPT (Baker et al., 2019) and we drew from the refined programme theory to design the RCT which is currently taking place. We anticipate that findings from the RCT and process evaluation will inform further refinement, representing

Figure 2. Graphic representation of the PARTNERS2 IPT.



overall a continuing, gradual and iterative process across the project that may potentially continue should other researchers draw from the theory in future.

Initial theory-building was conducted by the same members of the PARTNERS2 team later conducting formative and then process evaluation, although we purposely integrated discussion about development of the IPT content across the wider programme evaluation team during the initial theory-building stage. One researcher (RGJ) acted as 'the keeper of the theory'; she supported coherence by involvement with, communication about and integration of the different IPT sources and activities. Critical evaluation and questioning of proposed theory was encouraged within the team and with stakeholders as an approach to minimizing bias. We view the in-depth familiarity with the intervention content across the team as a strength that later supported particularly relevant formative and process evaluation designs, though for the full RCT we separate trial and process evaluation analysis.

In the sections below we describe the range of sources that we drew from to develop the IPT, only some of which had been specified in our original protocol. We then describe four activities which the PARTNERS2 team engaged in: articulation drawing from realist approaches; debate and consensus; communication; and interrogation. While the formal data sources contributing to the synthesis were planned in advance, the four activities contributing to synthesis represent our post hoc understanding of how we developed the model from these and knowledge held by stakeholders (service users, primary and secondary practitioners and

policy makers). Figure 3 depicts the overall process of IPT development. Figure 4 shows sources and activities. Although a collective understanding of the IPT was complete by October 2015, we continued to articulate and interrogate it after this time. Due to space constraints, in this paper we are only able to give an overview of methods in relation to data sources (additional detail can be requested from the authors).

Description of IPT sources

Our IPT was developed from two types of sources: formal data sources; and researcher and stakeholder knowledge and experience.

Formal data sources

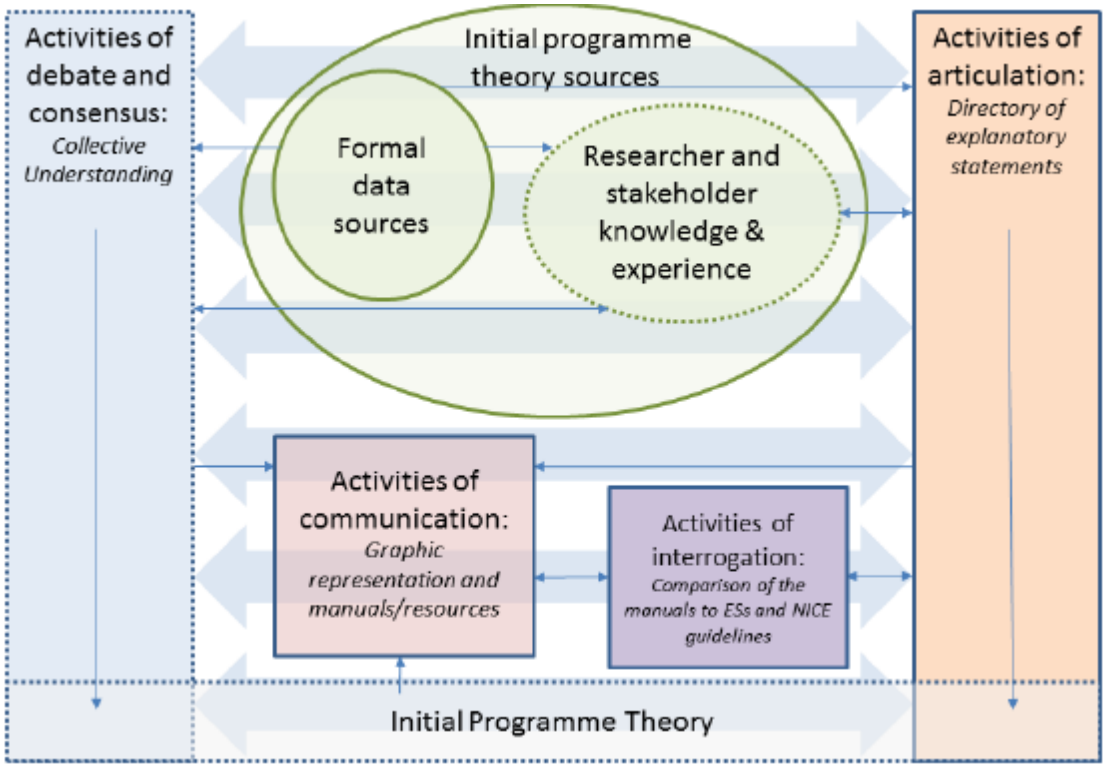
Formal data sources included literature on collaborative care and personal recovery, twelve interviews with key leaders in collaborative care and personal recovery, and six focus groups with service users. Ethical consent was granted by NRES Committee West Midlands – Edgbaston (REC reference number 14/WM/0052).

Research literature on collaborative care and personal recovery provided the foundational structure for components and key content of the IPT. Social science theory, including the Chronic Care Model (Wagner et al., 1996) and a conceptual framework for personal recovery in mental health (Leamy et al., 2011), with research evidence about collaborative care (Reilly S, 2013; Druss et al., 2001; Bauer et al., 2006; Kilbourne, 2008; van der Voort et al., 2015; Waxmonsky et al., 2014; Meadows et al., 2007; Chatterjee et al., 2014) supported our aim to follow fundamental principles of collaborative care optimised for people with experience of psychosis, where there was evidence of potential benefit.

Additional literature on personal recovery was surveyed and selected to represent views of service users, practitioners and policy makers (Bird et al., 2014; Bora et al., 2010; Brown and Kandirikirira, 2007; CSIP et al., 2007) about how and why personal recovery approaches were beneficial. One hundred and forty-four explanatory statements (ESs – see the section describing "articulation drawing from realist approaches" below) were written from this literature.

Eleven key experts were interviewed in order to explore their experience of how and why collaborative care and personal recovery approaches work. Because we drew from realist methods to articulate the IPT (see section below 'Articulation drawing from realist approaches'), it was agreed in keeping with realist principles of building on prior theory to formally draw from expertise on collaborative care in the researcher team, so RB and LG were invited to interview. Those who agreed (total 11; collaborative care: 10 researchers from the US (5), UK (3), Australia (1) and the Netherlands (1); personal recovery: 1 researcher from the UK) were emailed information about the study, details of our preliminary ideas for model components, and documents for obtaining their consent. Where consent was given, we interviewed these experts between October 2014 and February 2015 by telephone (9) or in person (2) about their experiences of intervention approaches, exploring how and why they

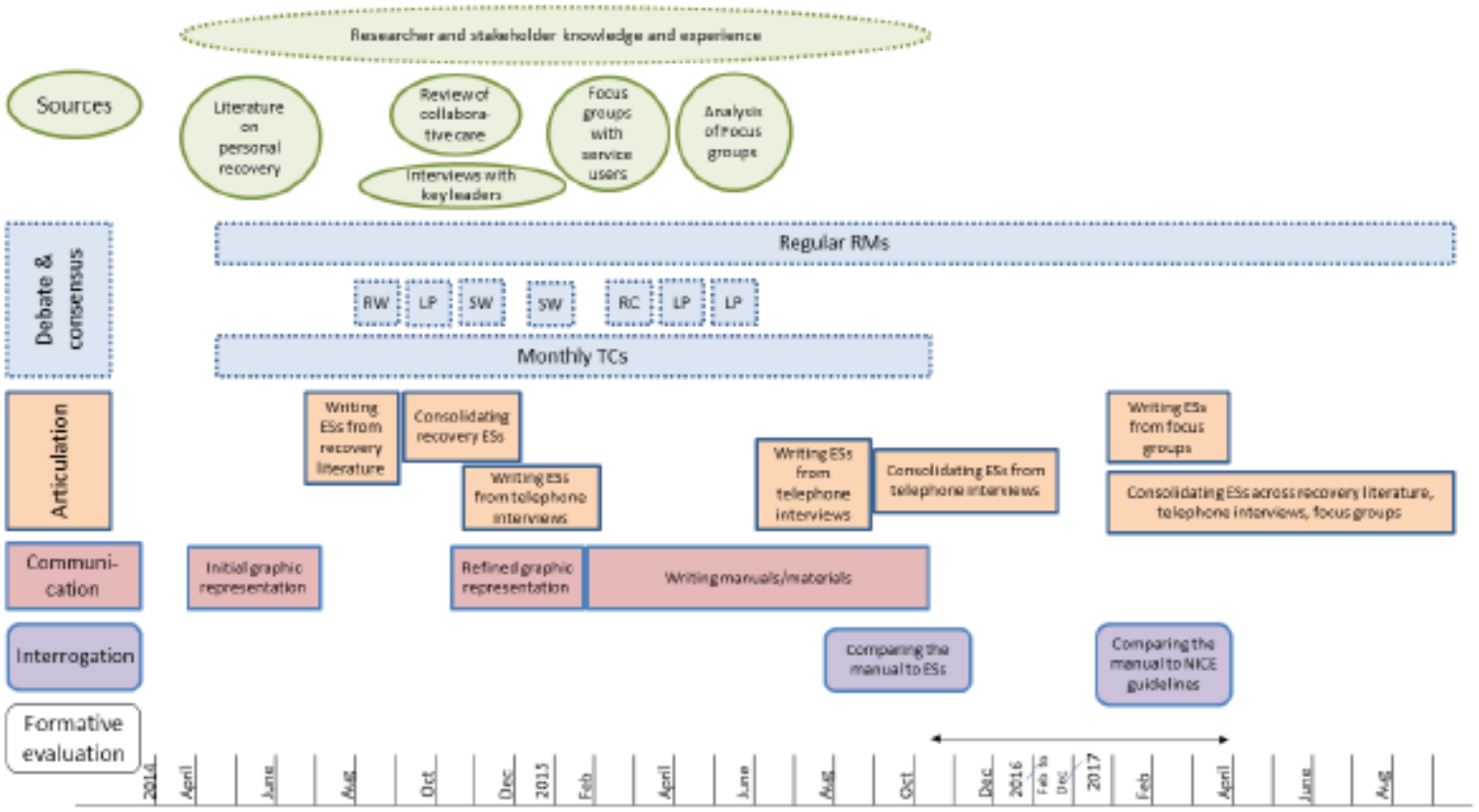
Figure 3. Diagram depicting the process of IPT development. Representations of the Model are shown with rectangles; sources of the IPT are shown with ovals; explicit elements have solid boundaries; more implicit elements have dotted boundaries. Outlines and arrows in blue represent the overlapping role of debate and consensus with other activities.



thought the PARTNERS2 intervention might or might not work. Interviews lasted 30-60 minutes. Two hundred and nine ESs were written from this data.

Focus groups with service users were held to explore current experiences of care. These were jointly held by two to three researchers from the PARTNERS2 team but were led by researchers with lived experience of mental health services. Service users who were not currently in crisis and had received care for psychosis in the previous two years were recruited through third sector organisations. At each of the three sites of the study, 1 focus group was held with participants with a diagnosis of schizophrenia (SZ) and 1 was held with participants with a diagnosis of bipolar (BP). A total of six focus groups involving 33 participants (13 women, 20 men) were conducted between January and March 2015, in Devon, Birmingham and Lancashire. Participants gave written consent, travel expenses were paid and each participant received a £10 gift voucher. Focus groups were audio recorded and transcribed. Transcriptions were coded in qualitative coding software (*Nvivo* 10), to collate data about processes of care, positive and negative experiences of care, and recommendations for care. Eighty nine ESs were written from this data.

Figure 4. Timeline showing sources and activities. **Key** ESs: Explanatory statements; LP: LEAP meeting; RC: researcher consensus meeting; RM: face to face researcher meeting; RW: researcher workshop; SW: stakeholder workshop; TC: researcher telephone conference. Representations of the IPT are shown with rectangles; sources of the IPT are shown with ovals; explicit elements have solid boundaries; more implicit elements have dotted boundaries.



Researcher and stakeholder knowledge and experience

Although researcher and stakeholder knowledge and experience was usually a less explicit basis for developing the IPT compared to the formal data sources, it provided a further source by contextualising, shaping and providing a referent from which to prioritise and evaluate the other data sources, as well as informing the writing of ESs.

Researchers, for example, brought additional ideas rather than just acting as 'neutral' programme theory builders. A number of researchers were purposely recruited to the project because they had experience of receiving mental health services, and they contributed expertise across the initial theory-building phase, but particularly during focus groups with service users, analysis of focus group data and writing and providing feedback about explanatory statements from focus groups and personal recovery literature. Clinician researchers each brought relevant knowledge and experience about collaborative care and methodology. For example, LG brought particular experience about supervision in collaborative care from her experience with other projects (Coventry et al., 2015; Richards et al., 2013). RB was involved in a separate collaborative care intervention for offenders (Lennox et al., 2018) which involved a realist review and evidence synthesis (Pearson et al., 2015). We also drew from the collaborative care expertise of LG and RB more formally by interviewing them as key leaders (see section above 'Formal data sources').

We recruited Lived Experience Advisory Panels (LEAPs) that met four times a year from third sector organisations in each of the three study sites. Meetings rotated by site so researchers had one LEAP to consult every month of the year for advice and guidance. The main role of the LEAPs were to provide expert input into the PARTNERS2 research study based upon experiential knowledge of psychosis, schizophrenia or bipolar. We recruited both service users and family members and each LEAP had up to 10 members. In each site, potential members attended an information gathering meeting in 2014 to assess what was involved and meet staff as well as other potential members. Criteria to join a LEAP were: interest in mental health research; understanding of secondary mental health and primary care services for people who experience psychosis and/or had a diagnosis of schizophrenia or bipolar; ideally experience of previous committee membership. We sought to build the panels with diverse perspectives through variations in gender, age, ethnicity and diagnostic label. The LEAP members gave feedback about explanatory statements, focus group topic guides and manual wording and content.

Description of activities to develop the IPT

Our IPT was developed through four activities: articulation drawing from realist approaches, debate and consensus, communication and interrogation. Realist approaches specifically guided the activity of articulation, however, we articulated in depth a number of issues around implementation, and the other activities we engaged in to create our IPT – debate and consensus, communication and interrogation – are not constrained by realist approaches and are relevant to theory-building processes more generally.

Articulation drawing from realist approaches

The activity of articulation involved identifying potential theories proposed within the formal data sources described above. Because realist approaches (Pawson, 2006; Pawson and Tilley, 1997) draw from a generative theory of causation that accounts for complexity and change over time in open systems, it can be a useful approach when considering complex interventions, so we adopted a realist approach to articulate our IPT. From a realist perspective, change due to an intervention does not only follow from the addition of intervention resources to a context, but is contingent on internal decisions by individuals. Such reasoned responses by individuals to intervention resources in a specific context is the mechanism by which an intervention brings about its outcomes (Lacouture et al., 2015). Exploration of the potential range of relational patterns between context, mechanisms and outcomes creates detailed programme theory, predicting how, why and for whom an intervention works. We also applied this approach to issues around implementation, where we probed stakeholders and the literature to identify theories about how the intervention activities could be introduced and maintained successfully. Realist approaches create mid-level theory (Pawson, 2010) that evaluators can draw from and adapt for use in other interventions that share similar aims.

The sources from which we drew to articulate the IPT include existing social science theory and medical research to identify broad structures and evidenced components, and we synthesised this with literature and stakeholder/researcher-level expertise to further elaborate and explain the intervention (for more detail of sources see below, Table 1). As is an aim within realist approaches, we were able to draw from relevant aspects of the programme theory of a project running in parallel, but ahead of, the PARTNERS2 project, the ENGAGER project (http://clahrc-peninsula.nihr.ac.uk/research/engager). Overlap of a co-applicant investigator (RB) between PARTNERS2 and ENGAGER bolstered this sharing.

Our methods for articulating the IPT involved adaptation of an approach to realist review adopted in the ENGAGER project (Pearson et al., 2015). We wrote explanatory statements (ESs) in the format 'If… then…', for example, "*If* the care partner acts as a three-way liaison, a conveyor of information between service user, general practitioner & community mental health team, *then* communication improves" [ES162]. Each ES describes a potential causal relationship within the intervention. We numbered the statements, recorded the data source and type of stakeholder, then categorised and consolidated ESs. We further drew from the ENGAGER project by applying the macro, meso and micro levels they identified (Pearson et al., 2015) as an initial framework for consolidation. We adapted these to be relevant to our data, identifying 9 categories: 1. Practitioner organisational, social and cultural context (macro); 2. Practitioner—practitioner interactions (meso); 3. Practitioner engagement and acceptability (meso); 4. Practitioner perceptions, understanding and skills (micro); 5. Service user—practitioner interactions (meso); 6. Service user experiences of care (meso); 7. Service user perceptions, understanding, skills and mental/physical health (micro); 8. Research aspects (intervention content, trial practicalities and approach to fidelity/process evaluation -- macro); 9. Carer perceptions, understanding and skills (micro). Researcher teams met repeatedly to discuss ES writing and consolidation. The directory of PARTNERS2 Consolidated ESs can be found in Supplementary figure S1.

Table 1. Contribution of each formal data source to the IPT.

Formal data source	**Contribution**	**Notes**
Social science theory	Source for foundational theory underpinning the Initial Model	(Leamy et al., 2011; Wagner et al., 1996)
Systematic review and/or studies on collaborative care for psychosis	Provided evidence and detail about foundational components for the Initial Model	(Bauer et al., 2006; Chatterjee et al., 2014; Druss et al., 2001; Kilbourne, 2008; Meadows et al., 2007; Reilly S, 2013; van der Voort et al., 2015; Waxmonsky et al., 2014)
Personal recovery literature	ESs* detailing content and approach to implementation at primarily meso and micro levels	Contributed mostly to ES* categories: 5. Service user – practitioner interactions; 7. Service users' perceptions, understanding, skills and/or physical and mental health.
Interviews with key leaders	ESs* detailing content and approach to implementation at primarily macro and meso levels	Contributed mostly to ES* categories: 1. Organisational, social and cultural context; 2. Practitioner—practitioner interactions; 3. Participant engagement/acceptability; 4. Practitioner perceptions, understanding and skills; 5. Practitioner—service user interactions; 8. Research aspects.
Focus groups with service users	ESs* detailing content and approach to implementation primarily at the meso and micro level	Contributed mostly to ES* categories: 6. Service users' experience of care 7. Service users' perceptions, understanding, skills and/or physical and mental health. Current perspectives of local service users substantiated literature on personal recovery.

*ES: Explanatory statement

Debate and consensus

While activities of articulation identified potential content, issues to consider around implementation and potential causal pathways for the PARTNERS2 IPT, the activities of debate and consensus between researchers and stakeholders determined the detail of what was and was not included in the IPT, and how it would be implemented.

Activities of debate and consensus were conducted during regular researcher telephone conferences, consultation with our LEAPs, stakeholder workshops where researchers, LEAP members and practitioners met, and face to face researcher meetings (see the timeline in Figure 4). The PARTNERS2 research team included a number of co-applicant investigators (RB, MB, LG, SR, NB, VP) with a range of specialisations including primary care, secondary care (clinical psychology and psychiatry), collaborative care, qualitative research and Patient & Public Involvement. Research Fellows, Associate Research Fellows, Service User Researchers and Researcher Consultants (RGJ, EB, HP, HL, JA, LGi, TR, JG, MC) also varied in their experience and knowledge of methodologies and mental health. The research team spanned seven universities (University of Birmingham, Lancaster University, University of Manchester, University of Warwick, University of Exeter, University of Plymouth and the London School of Economics and Political Science) and a research centre in London (The McPin Foundation). Varied areas of speciality and the distributed nature of the team meant that face to face meetings and workshops were particularly important, though limited for practical and financial reasons.

We drew from the knowledge and experience of stakeholders by consulting with LEAP service users and carers, primary and secondary health and mental health practitioners, and health commissioners recruited from our three intervention sites in Lancashire, Birmingham and Devon.

Finally, processes of discussion and negotiation with local Trusts circumscribed or shaped the content of the IPT. In some instances, we were unable to adopt aspects of the intervention because of local research site contexts. For example, personal recovery literature, the interview with a key leader in personal recovery, focus group data and input from stakeholder workshops suggested employing peer support workers to carry out the care partner role to support egalitarian relationships. However, there were not adequate numbers of peer support workers within the local Trusts at our research sites to enable this. Local contexts therefore contributed to initial theory-building.

Communication

Aspects of the IPT were made explicit in order to communicate PARTNERS2 to others through a graphic representation (see Figure 2, the PARTNERS2 initial model) and manuals (available on request to the authors).

The graphic representation is a visual summary of the PARTNERS2 intervention, and highlights the main types of intervention resources that include structures, information and people. Change occurs primarily at two levels; practitioner change following training, resources and supervision, and service user change following interactions with a care partner including engagement and retention, coaching, care coordination and review. We also communicated the IPT by writing manuals for care partners and their supervisors, service

users, carers and GPs. We adapted some of the resources from the ENGAGER intervention (Pearson et al., 2015) and part of its manual framework and content where its aims were similar to the aims of PARTNERS2. The manuals were initially written by a few researchers based on the collective understanding, and subsequently debated across the research team and LEAPs, and revised at length.

Interrogation

We interrogated the IPT internally by comparing content between different representations of the intervention, and externally by comparing the practitioner manual to clinical guidelines for best practice. The practitioner manual was compared to the directory of 442 ESs, in order to check consistency and identify gaps. The two representations of the model were highly consistent, although a few gaps, where the manual did not completely represent ESs were identified. For example, responsibilities of the care partner and supervisors in relation to liaison with primary care staff needed further clarification. These gaps were discussed across the researcher team, and the manual was adapted to more fully reflect the ESs.

The manual was compared to relevant NICE clinical guidelines (NICE, 2006; NICE, 2012; NICE, 2014) in order to explore how consistent it was with current guidelines for good practice, and to identify any gaps. The comparison showed the manual and guidelines to be mostly consistent. The gaps identified related to issues that had been agreed within the research team, but that had been postponed due to lack of capacity (e.g. creating a Carers/Friends and Family manual and directories of local resources; manualising preparation for the end of PARTNERS2) or were identified in the ESs but not the manual (e.g. being sensitive to service users' multiple identities). These gaps were discussed across the research team, and we adapted our ESs and/or the manual to address them.

Lessons learnt about developing the IPT

The four activities described above worked interactively in a non-linear manner. Activities of debate and consensus evaluated, circumscribed, structured and/or guided the impacts of articulation, communication and interrogation. Articulation provided evidenced content for debate and consensus. Activities of communication created public representations of the IPT built from processes of articulation, debate and consensus, and interrogation. Interrogation established consistency and robustness of the processes of articulation, debate and consensus, and communication helped reduce the risks of bias from any one source.

Through ongoing dialogue about the intervention, a collective and increasingly explicit understanding was created between PARTNERS2 researchers and LEAP members about what the intervention involved. In an ongoing iterative process, all four activities both contributed to, and were tailored by, the developing collective understanding. Inevitably individual researchers and LEAP members understood the model in slightly different ways, but the collective understanding represented distributed meanings across the PARTNERS2 team.

Below we reflect on each of the activities in turn, then discuss a particularly beneficial aspect of the interaction between activities of articulation, and debate and consensus.

Lessons learnt about articulation

The activities of articulation created a foundation and process for establishing content of the IPT. From social science theory (Wagner et al., 1996; Leamy et al., 2011) we established core components of the model. The PARTNERS2 directory of ESs explicated in detail a number of potential causal patterns that provide a transparent record of explanation for the content of the intervention and how we anticipated it would work. Organising ESs into macro-meso-micro levels clarified content at different levels within the intervention, and supported understanding about relationships between them.

Different formal data sources provided detail for different categories and levels of ESs (see Table 1). This demonstrated how important it was to draw across sources chosen in order to illuminate different aspects of the complex PARTNERS2 intervention. Drawing from a range of sources was also beneficial because information from one source often provided support for or challenge to the relevance and meaning of issues flagged in other sources. Where sources supported each other, this substantiated potential ESs; where sources challenged each other, this highlighted areas requiring debate and consensus. Through these processes we were able to reduce the risks of bias from any one source.

There were also a number of difficulties related to ESs. Originally, our intention was for the directory of ESs to represent the IPT as fully as possible. However, because of project financial constraints and long term illness of more than one researcher, capacity was limited during the first three years of the study. The time needed for this process was already lengthy due to unfamiliarity with the role of realist methods in developing our IPT for most of the researcher team and its time-consuming nature. Our process of writing ESs was adapted as a result.

Initially, we intended to write ESs systematically not only in response to literature and primary data, but also researcher and stakeholder discussions, citing each meeting as a source. Due to time limitations, instead of writing ESs following discussions in meetings, during identification and consolidation of ESs we adapted the original text of data sources to be consistent with the content of the PARTNERS2 collective understanding. We also intended to refine ESs further, for example by consolidating the individual-level service user theory around personal recovery by mechanism as well as outcome, because the relationships tended to be bi-directional. This process was curtailed, along with intentions to write narratives for each of the nine main categories of theory, due to time constraints. Even with these adaptations, the process of identifying and consolidating ESs continued on through the formative evaluation. In effect, the collective understanding, graphic representation and manuals represented our IPT that was tested during formative evaluation, while the ESs were more slowly explicated. Eventually we moved on to the needs of conducting the process evaluation of the upcoming RCT, rather than developing the ESs further. The time-consuming nature of the process of identifying and consolidating ESs was problematic, though it provided a robust basis for developing our IPT.

Lessons learnt about debate and consensus

A particular benefit to the intervention resulting from activities of debate and consensus was the identification of, and work to prevent, unintended consequences. For example, there was

an awareness from the focus groups with service users that in changing healthcare providers (the intervention requires relocating service users' care from secondary mental health care services to primary care) there was the risk we would destabilise existing, supportive relationships between practitioners and service users. In addition, until agreements had been negotiated with local Trusts who provided the secondary mental health care, we did not know how service users would be re-integrated into existing mental health services when the PARTNERS2 intervention concluded, or if the collaborative care model might continue locally beyond the trial. Through ongoing consultation with the PARTNERS2 LEAPs, service users and carers reiterated transfer in and out of PARTNERS2 as a crucial aspect of the intervention, and we developed recruitment materials and intervention resources and content to support transfer back into usual care after leaving PARTNERS2. While the intervention's approaches to transfer have not yet been fully tested, the input from LEAPs enabled more nuanced, sensitive and complete resources to be developed.

Lessons learnt about communication

The graphic representation of the model was used during recruitment to explain the intervention to local Trusts and GP surgery staff, and was included in the Care Partner/Supervisor Manual. Although of perhaps limited use in isolation, it provided a shorthand for the contents of the intervention, and a reference for ongoing discussion about what the intervention involved.

The manuals were highly important during the lead up to the pilot phase and formative evaluation; they described how the intervention was to be carried out in practice. The content of the practitioner manuals included aspects of the model that we anticipated might not already be part of usual care, in order to focus on areas of change. Service user and Friends and Family manuals explained the support they would receive and clarified roles of practitioners and the service user, in order to inform and direct expectations. Because of the extended time period necessary to complete the directory of ESs and because of its complexity, the PARTNERS2 manuals acted as the primary means of communication about the IPT during the formative evaluation and after adaptation in the main trial.

Lessons learnt about interrogation

The activities of interrogation represented self-checking exercises for the IPT, both between different elements within the IPT and in comparison with external guidelines for good practice. They allowed us to systematically identify and remedy gaps. The consistency established by these two activities of interrogation supported our confidence in the rigour and quality of the IPT for PARTNERS2.

Interactions between articulation and debate and consensus

A beneficial but unanticipated aspect of activities of articulation involved their effect of specifying and grounding what might have otherwise been more abstract concepts, which facilitated activities of debate and consensus. The level of the ESs (relating to specific circumstances and issues) reduced perceptions that the theory was over-abstract or removed from practice and the experiences of service users and carers. The process of developing ESs therefore seemed to create a structure and bridge for researchers and stakeholders to

meaningfully move back and forth between practice and theory. The process of realist synthesis supported the development of IPT not only as a method for articulating causal patterns and synthesising these, but perhaps as importantly, by creating a structure and focus for negotiating understanding across researchers and stakeholders with widely divergent experience and knowledge.

Processes of data collection and analysis framed multiple potential aspects of the intervention with a focus on 'why' and 'how'. This brought to discussion many topics and initiated face to face meetings between researchers and stakeholders that might not have occurred otherwise, and created a context for discussion that naturally moved beyond 'this is what I think we should do', to 'we could do this because'. In other words, it encouraged focus on the reason *behind* choices that was less influenced by status, expert opinion and/or gestalt meanings.

One example involves two of the co-applicant investigators who were practitioner researchers from primary (RB) and secondary (LG) care and who had extensive experience working with collaborative care interventions (Lennox et al., 2018; Coventry et al., 2015; Richards et al., 2013). Their input carried particular weight in the theory-building process though their experiences and perspectives were quite different. RB tended to think and communicate on a more conceptually abstract basis, whereas LG tended to prioritise a more grounded, pragmatic approach. The two researchers' different styles, combined with different practitioner emphases, could sometimes create barriers to understanding and/or consensus between them. Initially, LG expressed reservations about framing collaborative care with principles of personal recovery, because its meaning could be misunderstood by practitioners and service users (Slade et al., 2014) and because it was often discussed in highly abstract terms. There was also concern across the team that inclusion of such principles mistakenly suggested the focus of collaborative care was as much about care partner—service user collaboration as about primary—secondary care collaboration. These differences were resolved through open discussion and debate as there were found to be few if any substantive differences of view once differences related to language had been understood.

Through activities of articulation such as reviewing research on collaborative care and psychosis, interviews with key leaders and writing and consolidating ESs from literature on personal recovery, we were able to identify implicit aspects of personal recovery in existing collaborative care interventions and convincing beneficial aspects of a personal recovery approach in relation to psychosis, which led to consensus. We operationalised these for practice through adoption of coaching principles (Bora et al., 2010) to guide interactions between the PARTNERS2 care partners and service users. Explicit incorporation of recovery principles and our approach to their operationalization is a distinguishing feature of PARTNERS2 collaborative care for people who experience psychosis.

Discussion

In this paper we have provided a worked example of the initial theory-building phase of TDE for PARTNERS2 collaborative care for psychosis, a complex intervention. We have described data sources and IPT development activities, and reflected on the lessons learnt. While not intended to be a guide, we hope that our post hoc 'warts and all' description of the four varied activities we ended up engaging in to create the IPT, and our discussion of the interactions between these activities, will be helpful to others developing complex

interventions. We drew from similar activities again during our formative evaluation (Baker et al., 2019), and anticipate these activities will support us in evaluating the intervention during our RCT and process evaluation. Although we only report the activities in detail for the initial theory-building phase, we posit that they are beneficial within evaluation phases of TDE as well.

We would now like to draw on our experience during this process to contribute to the debate on the value of developing IPT within a TDE approach. We return to the two objections (Coryn et al., 2011) to TDE raised in the Introduction to this paper, that 1) explication of programme theory is unnecessary because it is often not used in any meaningful way, and 2) since developing high quality programme theories is often not feasible, and poor quality programme theories can be counter-productive, conducting TDE is a waste of valuable resources. In contrast to the first objection, the process of IPT development contributed meaningfully to our complex intervention in a number of ways, including: a) detailed clarification of the intervention rationale; b) establishing a high level of internal consistency between IPT and communications of the model such as the intervention manuals; c) highlighting issues that prepared us to better train practitioners; d) creating a framework for thorough evaluation including hypotheses about why the intervention might or might not work; and e) providing a structure that fostered dialogue and understanding amongst researchers, service users, carers and practitioners, allowing more egalitarian and thorough exploration of the issues around collaborative care for people who experience psychosis.

With regards to the second objection, our experience both supported and refuted it in different ways. In accordance with the charge that high quality programme theory development is not feasible, we found the activities of articulation drawing from realist approaches as well as the activity of debate and consensus to be particularly time consuming, and we could easily have spent additional time in developing the IPT further. Other projects have reported similar difficulties (Lloyd et al., 2017). As recipients of a 5-year Programme Grant from the National Institute of Health Research in England, we were able to expend extensive resources on initial theory-building, but smaller research projects may not be able to do so, and this is a limitation of the approach. Drawing on a realist approach for part of this initial theory-building process helped bring a particular rigour, but it was resource intensive. In addition, we added to it the other activities to ensure a grounding across multiple stakeholders with actual experience of the phenomena of interest and a language that resonated with this, and spent time checking the IPT internally and externally. These activities were also intensive but, in our experience, invaluable elements to combine with the realist approach. Finding the right amounts of time and resources to develop a programme theory robust enough for the purposes of each project remains a matter of judgment and fine balance.

In refutation to the second objection, although we were obliged to follow Coryn's (2011) recommendation for the need to prioritise and balance programme theory development in the face of pragmatic limitations, we *were* able to develop an internally coherent complex intervention that is documented in detail. We found the realist approach of tabulating and consolidating hypothetical relationships, although time consuming, supported high quality activities of debate and consensus, and particularly in-depth conceptualisations about how the intervention might work, and for whom. And, having expended these resources, other researchers and healthcare practitioners can now draw from the study as a resource.

A further question concerning the programme theory at this time is that its utility and efficacy have yet to be tested in a full trial. A further strength is the large number of collaborators involved in the initial theory-building phase, which guards against individual biases or idiosyncrasies. We have contributed to the cumulation of research around collaborative care, by drawing from existing theory and strengthening this with primary data and further development. Because the IPT is mid-range, it is likely to be generalisable to other interventions with similar aims applied within similar contexts.

The work done on PARTNERS2 is timely. The Independent Mental Health Taskforce has recently published the five year forward view for mental health in the UK (Farmer and Dyer, 2016), which includes recommendations for improved physical healthcare for people with more severe mental health problems, and support for mental health from primary care. In future, UK Healthcare Trusts may therefore be likely to work on the implementation of models with aims similar to PARTNERS2. The IPT that we have built is deeply rooted in existing literature and theory on collaborative care and personal recovery as well as the experience of many experts and stakeholders. We hope others will develop it further and use this as a resource for evaluation work in the future.

Authors' contributions

RGJ, RS, HL and VP identified and prioritised personal recovery literature. RGJ, RB, NB, JA, LGi, HL and EB wrote explanatory statements. LGa, RB, RGJ, NB, VP and SR identified key leaders, conducted key leader interviews, read and commented on interview transcripts. LGa, RGJ, HP, VP and RS planned and facilitated stakeholder workshops. TR, LGi, VP, RGJ, JA, RS, HP, NB and SR planned and/or conducted focus groups and/or analysed focus group data. LGa, RGJ, RB, LGi, VP and RS wrote manuals representing the Initial Model. LGa, RB, RGJ and HP wrote and planned practitioner training and resources to communicate the Initial Model. JA and RGJ interrogated the Initial Model in collaboration with LGa, RB, SR and VP. RGJ wrote the draft manuscript, and NB, RB, LGa, VP, SR, LGi, RS, EB, JA, HL, JG, TR, MC and MB provided comments on the draft.

Acknowledgements

We would like to acknowledge Professor Helen Lester's substantial contribution to the development of this programme grant. Professor Lester conceived and wrote much of the original protocol for this work which was successfully funded as a programme grant in 2012. Professor Lester passed away on the 2nd March 2013.

We would also like to thank Maria Cox, Gemma Taylor and Julie Billsborough for facilitating focus groups and/or for the analysis of focus group data, Simon Love for his contributions to refinement of the personal recovery ESs and Terry Davies for his input into practitioner training. We would also like to thank the primary and secondary healthcare practitioners and commissioners who offered their expertise at stakeholder workshops. Finally, we would like to pay special tribute to each member of the three PARTNERS2 LEAPs in Devon, Lancashire and Birmingham for their valuable contributions across the initial theory-building phase.

Declaration of Conflicting Interest

The Authors declare that there is no conflict of interest.

Supplementary materials

Supplementary materials can be accessed from the corresponding author.

Funding acknowledgement

This research was funded by a UK NIHR Programme Grant (RP-PG-0611-20004) and the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula (NIHR CLAHRC South West Peninsula). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

References

Anderson R. (2008) New MRC guidance on evaluating complex interventions. Br Med J, 337:a 1937.

Baker, et al. (2019) Refining a model of collaborative care for people with a diagnosis of bipolar, schizophrenia or other psychoses in Engalnd: a qualitative formative evaluation. *BMC Psychiatry* 19(1):7.

Bauer M, McBride L, Williford W, et al. (2006) Cooperative Studies Program 430 Study Team.

Bird V, Leamy M, Tew J, et al. (2014) Fit for purpose? Validation of a conceptual framework for personal recovery with current mental health consumers. *Aust N Z J Psychiatry*.

Blamey A and MacKenzie M. (2007) Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges? *Evaluation* 13(4): 439-455.

Bora R, Leaning S, Moores A, et al. (2010) Life coaching for mental health recovery: the emerging practice of recovery coaching. *Advances in Psychiatric Treatment* 16: 459-467.

Brand S, Quinn C, Pearson M, et al. (2018) Building programme theory to develop more adaptable and scalable complex interventions: Realist formative process evaluation prior to full trial. *Evaluation*. Epub ahead of print DOI: 10.1177/1356389018802134.

Brown W and Kandirikirira N. (2007) Recovering mental health in Scotland. Report on narrative investigation of mental health recovery. Glasgow: Scottish Recovery Network.

Chatterjee S, Naik S, John S, et al. (2014) Effectiveness of a community-based intervention for people with schizophrenia and their caregivers in India (COPSI): a randomised controlled trial. *Lancet* 383: 1385-1394.

Chen HT. (1990) *Theory Driven Evaluation*, Thousand Oack, CA: Sage.

Collaborative care for bipolar disorder: part 1 intervention and implementation in a randomized effectiveness trial. *Psychiatric Services* 57: 927-936.

Connell JP and Kubisch A. (1995) Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiative. In: Connell JP (ed) *New Approaches to Evaluating Community Initiatives. Concepts, Methods, and Context. Roundtable on Comprehensive Community Initiatives for Children and Families*. Washington DC: The Aspen Institute.

Coryn CLS, Noakes LA, Westine CD, et al. (2011) A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. *American Journal of Evaluation* 32: 199-226.

Coventry P, Lovell K, Dickens C, et al. (2015) Integrated primary care for patients with mental and physical multimorbidity: cluster randomised controlled trial of collaborative care for patients with depression comorbid with diabetes or cardiovascular disease. *BMJ : British Medical Journal* 350.

Craig P, Dieppe P, Macintyre S, et al. (2006) Developing and evaluating complex interventions: new guidance. Medical Research Council.

Craig P, Dieppe P, Macintyre S, et al. (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 337.

CSIP, RCPsych and SCIE. (2007) A common purpose: Recovery in future mental health services. London: SCIE (Social Care Institute for Excellence).

De Silva MJ, Breuer E, Lee L, et al. (2014) Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials* 15: 267.

Druss B, Rohrbaugh R, Levinson C, et al. (2001) Integrated medical care for patients with serious psychiatric illness: a randomized trial. *Archives of General Psychiatry* 58: 861-868.

Farmer P and Dyer J. (2016) The Five Year Forward View for Mental Health. The Independent Mental Health Taskforce to the NHS in England.

Funnell SC and Rogers PJ. (2011) *Purposeful Program Theory: Effective use of theories of change and logic models,* San Francisco, CA: Jossey-Bass.

Hawe P, Shiell A and Riley T. (2009) Theorising Interventions as Events in Systems. *Am J Community Psychol* 43: 267-276.

Kilbourne AM. (2008) Improving medical and psychiatric outcomes among individuals with bipolar disorder: a randomized controlled trial. *Psychiatr Serv* 59: 760-768.

Kirkpatrick DL. (1959) Techniques for Evaluating Training Programmes. *J Am Soc for Train Dev* 11: 1-13.

Lacouture A, Breton E, Guichard A, et al. (2015) The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. *Implementation Science* 10.

Leamy M, Bird V, Le Boutillier C, et al. (2011) Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *The British Journal of Psychiatry* 199: 445-452.

Leeuw FL and Donaldson SI. (2015) Theory in evaluation: Reducing confusion and encouraging debate. *Evaluation* 21: 467-480.

Lennox C, Kirkpatrick T, Taylor R, et al. (2018) Pilot randomised controlled trial of the ENGAGER collaborative care intervention for prisoners with common mental health problems, near to and after release. *Pilot and Feasibility Studies* 4.

Lloyd HM, Pearson M, Sheaff R, et al. (2017) Collaborative action for person-centred coordinated care (P3C): an approach to support the development of a comprehensive system-wide solution to fragmented care. *Health Research Policy and Systems* 15: 98.

Meadows GN, Harvey C, Joubert L, et al. (2007) The Consultation-Liaison in Primary-Care Psychiatry Program: A structured approach to long-term collaboration. *Psychiatr Serv* 58: 1036-1038.

Moore GF, Audrey S, Barker M, et al. (2015) Process evaluation of complex interventions. UK Medical Research Council.

Moore GF and Evans RE. (2017) What theory, for whom and in which context? Reflections on the application of theory in the development and evaluation fof complex population health interventions. *SSM - Population Health* 3: 132-135.

NICE. (2006) Bipolar Disorder: the Management of Bipolar Disorder in Adults, Children and Adolescents, in Primary and Secondary Care. National Institute for Health and Care Excellence.

NICE. (2012) Service User Experience in Adult Mental Health: Improving the Experience of Care for People using Adult NHS Mental Health Services. National Institute for Health and Care Excellence.

NICE. (2014) Psychosis and Schizophrenia in Adults: The NICE Guideline on Treatment and Management. National Institute for Health and Care Excellence.

Pawson R. (2006) *Evidence-based Policy: A Realist Perspective,* London: Sage.

Pawson R. (2010) Middle range theory and program theory evaluation: From provenance to practice. In: Vaessen J and Leeuw FL (eds) *Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences.* New Brunswick, NJ: Transaction Publishers, 171-202.

Pawson R and Tilley N. (1997) *Realistic Evaluation,* London: Sage Publications.

Pearson M, Brand SL, Quinn C, et al. (2015) Using realist review to inform intervention development: methodological illustration and conceptual platform for collaborative care in offender mental health. *Implement Sci* 10.

Reilly S PC, Gask L, Hann M, Knowles S, Druss B, Lester H. (2013) Collaborative care approaches for people with severe mental illness (Review). *The Cochrane Library*.

Richards DA, Hill JJ, Gask L, et al. (2013) Clinical effectiveness of collaborative care for depression in UK primary care (CADET): cluster randomised controlled trial. *BMJ : British Medical Journal* 347.

Shearn K, Allmark, P Piercy H, et al. (2017) Building Realist Program Theory for Large Complex and Messy Interventions. *International Journal of Qualitative Methods* 16: 1-11.

Slade M, Amering M, Farkas M, et al. (2014) Uses and abuses of recovery: implementing recovery-oriented practices in mental health systems. *World Psychiatry* 13: 12-20.

Stame N. (2004) Theory-based Evaluation and Types of Complexity. *Evaluation* 10(1): 58-76.

Suchman E. (1967) *Evaluative Research*. New York, NY: Russell Sage Foundation.

van der Voort TYG, van Meijel B, Goossens PJJ, et al. (2015) Collaborative care for patients with bipolar disorder: randomised controlled trial. *The British Journal of Psychiatry* 206: 393-400.

Wagner EH, Austin BT and Von Korff M. (1996) Organizing Care for Patients with Chronic Illness. *The Milbank Quarterly* 74: 511-544.

Waxmonsky J, Kilbourne AM, Goodrich DE, et al. (2014) Enhanced Fidelity to Treatment for Bipolar Disorder: Results From a Randomized Controlled Implementation Trial. *Psychiatr Serv* 65: 81-90.

Weiss CH. (1995) Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In: Connell JP (ed) *New Approaches to Evaluating Community Initiatives. Concepts, Methods, and Context. Roundtable on Comprehensive Community Initiatives for Children and Families.* Queenstown, MD: Aspen Institute Publications Office.

Weiss CH. (1997) Theory-based Evaluation: Past, Present and Future. In: Rog DJ (ed) *Progress and Future Directions in Evaluation.* San Francisco, CA: Jossey-Bass.