

**Fidelity measurement for the implementation of social networks
interventions in complex mental health**

Mauricio Alvarez Monjarás

D.Clin.Psy. thesis (Volume 1), 2019

University College London

Overview

This thesis examines the concept of programme fidelity as related to multi-component mental health interventions for severe mental health difficulties.

Part 1, the literature review, explores existing measures in the field to identify the ‘best standards’ for programme fidelity assessment and an initial articulation of their typology. 31 papers are examined using a narrative synthesis approach. The content and psychometric properties of the 12 identified measures are described, identifying common practices, and highlighting neglected assessment areas. Findings highlight a scarcity of measures despite their increasing relevance to the field of implementation science. Further, although arguably consistent in design, assessment procedures, and scoring, there is great variability in the domains covered.

Part 2, the empirical paper, describes the development, piloting and implementation of the Community Mental Health Fidelity Scale (CoMFideS), as part of the Open Dialogue: Development and Evaluation of a Social Network Intervention for Severe Mental Illness (ODDESSI) trial. The ODDESSI trial aims to assess whether Open Dialogue (OD) –when integrated within standard NHS mental health services for adults in crisis– improves the clinical and cost-effectiveness of traditional crisis and continuing community mental health care (CAU). CoMFideS was developed based on findings from the systematic review and drawing from expertise and existing measures in the field. CoMFideS was piloted in 6 OD and 6 CAU services. Findings suggests that CoMFideS may be a reliable and feasible programme fidelity measure of high-quality CAU and specific OD features.

Part 3, the critical appraisal, provides some final thoughts about the research project, including a discussion on some methodological and theoretical considerations, as well as some reflections on the research process as a whole.

Impact Statement

The present research has a series of academic, clinical, and service level implications beyond its use for the Open Dialogue: Development and Evaluation of a Social Network Intervention for Severe Mental Illness (ODDESSI) trial.

On an academic level, fidelity assessment is an important –yet somewhat neglected– aspect of implementation science. The literature frequently reports a scarcity of validated measures and a lack of consistency in the domains of service provision that need to be assessed. Here, by designing an exhaustive and elaborate systematic search strategy, an extended and detailed list of validated programme fidelity measures of multicomponent interventions for severe mental illness is presented. Further, our in-depth analysis offers a series of key fidelity domains that could guide the development of new measures, as illustrated in our empirical study. Results from our systematic review might, therefore, provide insights for the typology of programme fidelity measures and help move toward a more systematic approach to fidelity assessment.

This study also extends our understanding of the role of fidelity assessment in cross-cultural adaptations of psychosocial interventions, in this case, the Open Dialogue approach. By using recognised and recommended methods of measure development we demonstrate that it is possible to translate abstract theoretical and clinical principles into measurable components of service delivery. The Community Mental Health Fidelity Tool (CoMFideS) was able to identify some structural and functional features of the Open Dialogue approach that may be transferrable to the National Health Service and provides some initial insights as to which components of the model may not be structurally identifiable. Future research may help establish the psychometric properties and internal structure of the CoMFideS and, potentially,

assess its applicability in other community mental health settings or other health systems across the globe.

The clinical implications of this research are also worth noting. The CoMFideS uncovered new venues of research in dialogic research and practice. To date, most of the literature focuses on network meetings as the key component of dialogic practices; however, little attention is paid to the structures around these meetings. Here, we take a step outside the network meetings and question what kind of contextual and structural variables may make Open Dialogue a unique approach to crisis and continued community care. It is yet to be explored which of the structural domains here identified have a greater influence on outcomes or whether they are in any way related to other clinically-relevant aspects of the approach, such as therapist adherence or service-user involvement. Hopefully, results from the ODDESSI trial will help elucidate these doubts and help enhance pieces or training in dialogic practice and standard crisis and continuing community care.

Finally, and perhaps most importantly, this research offers a reliable method for understanding the inner workings of community mental health settings that is worthy of dissemination. As a research tool CoMFideS may be used as a means to ensure that services participating in trials have similar standards of care and are therefore comparable to each other. As a self-monitoring tool, CoMFideS can be helpful for service managers and clinicians to identify strengths and areas in need of training or resources. Whichever the case, this study presents a novel resource for the Open Dialogue community and community mental health services in general to appraise their social networks orientation.

DClinPsy Thesis (Volume 1): Major Research Project

Table of Contents

Acknowledgements	9
Part 1. Literature review	11
Abstract	12
Introduction	13
Method	21
Results	25
Discussion.....	40
References	47
Part 2. Empirical paper	61
Abstract	62
Introduction.....	63
Method	69
Results	83
Discussion	91
References	102
Part 3. Critical appraisal	112
Overview	113
Methodological considerations	113
Defining the context	115
Dialogue and power	117
The value of anthropology	118
Final reflections: The scientist practitioner stance	120
References	121

Appendix A	123
Appendix B	126
Appendix C	128
Appendix D	130
Appendix E	134
Appendix F	139
Appendix G	145
Appendix H	148
Appendix I	152
Appendix J	161
Appendix K	209
Appendix L	211

Index of Tables

Literature review

<i>Table 1.</i> Programme fidelity measures	26
<i>Table 2.</i> Common domains across measures	35
<i>Table 3.</i> Process features by measure	36
<i>Table 4.</i> Organizational structure features by measure	37
<i>Table 5.</i> Delivery of care features by measure	38

Empirical paper

<i>Table 1.</i> Community Mental Health Fidelity Scale items	75
<i>Table 2.</i> Service characteristics	84
<i>Table 3.</i> Differences in CoMFideS mean scores between respondents	85
<i>Table 4.</i> Inter-rater reliability of the CoMFideS using the G estimate	86
<i>Table 5.</i> Internal consistency of the CoMFideS	87
<i>Table 6.</i> Differences in CoMFideS mean scores between service models	87
<i>Table 7.</i> Lowest and highest scoring items across teams	88
<i>Table 8.</i> CoMFideS mean scores across teams	91

Index of Figures

<i>Figure 1.</i> Community Mental health Fidelity Scale development process	71
<i>Figure 2.</i> Comparison of mean CoMFideS section scores between OD and CAU..	88
<i>Figure 3.</i> Comparison of mean CoMFideS item scores between OD and CAU	89
<i>Figure 4.</i> Mean scores of the Open Dialogue addendum	90

Acknowledgements

I would like to give my gratitude to the Mexican Council of Science and Technology (CONACYT) for their financial support throughout this doctoral degree. Without them, I would have never been able to accomplish this goal. I hope the learning and experience gained thus far eventually have an impact in my country.

My deepest gratitude and admiration to my supervisor, Professor Steve Pilling. Thank you for trusting in this project and for your academic and professional mentorship; your advice, support, and experience were vital to this study. I hope CoMFideS serves its purpose as a useful tool for the ODDESSI trial and contributes to the advancement of these and other community mental health interventions.

I would also like to thank the ODDESSI 'Power House': Emily Wilson, Katherine Clarke, and Melissa Lotmore. Without their immense support, their patience with my never-ending emails, and their help in conducting the fidelity interviews I would not have been able to bring this project to a close. Thanks as well to Christopher Cooper for helping us consolidate our search strategy and thanks to all ODDESSI research assistants and members of the steering committee for arranging the interviews and for sharing your insights about CoMFideS. Most importantly, thank you to all the Open Dialogue and CAU managers and practitioners, without whom this project would not have been possible.

I would also like to thank Dr Russell Razzaque and Mark Hopfenbeck for their support and advice throughout the project; their knowledge on Open Dialogue was key for the development of the addendum. I hope that CoMFideS proves to be fruitful for the trial and supports the dissemination of Open Dialogue in United Kingdom and abroad. Special thanks as well to Dr Lauren Cubellis and Dr Inga-Britt Krause who ignited my passion for anthropology by revealing the value of the

discipline to the field of mental health.

A very big thank you goes to the UCL 2016-19 DClinPsy cohort for their friendship and for making me feel at home since day one. Thanks to the UCL Department of Clinical, Educational, and Health Psychology for the teaching and for walking alongside us throughout this adventure. Thank you to my tutor, Dr Miriam Fornells-Ambrojo, for her kindness and patience with all my scholarship paperwork and training needs. Thanks as well to Dr Joshua Stott for his thoughtful MPR reviews and for ensuring I made the most of my clinical placements. A very big thanks to all my clinical supervisors (Allison Wilson-Thompson, Dr Meenaxi Patel, Dr Barry Star, Adriana Fernandez-Chirre, Dr Sian Granville, and Dr Ollie Hawthorne) for your reflexivity, honesty, and dedication, and for helping me find (and own) my therapeutic voice as I developed my identity as a clinician.

To my friends, I thank them for their unfaltering support, their patience, and for helping me keep my eyes on the goal.

Finally, the biggest thank you goes to my family: Monica, Alfredo, and Bernardo. Despite the distance, your love and encouragement have kept me going all these years. I dedicate this thesis to you.

Thank you for believing in me.

Part 1: Literature Review

Programme fidelity measures of complex mental health interventions for
severe mental health difficulties: A systematic review

Abstract

Programme fidelity assessment is an important component of an intervention's implementation evaluation and a service's quality assurance, given its strong correlation with client and service outcomes. It is a key factor in ensuring internal and external validity of an intervention as it helps clearly specify its delivery parameters and differentiate them from other interventions. Unfortunately, programme fidelity assessment is uncommon in psychosocial implementation studies, particularly in complex or multi-component interventions. This is partly due to a lack of clarity of the terminology, an uncertainty concerning its key domains, and the level of specificity required. Further, assessing these interventions effectively and efficiently can be challenging given the substantial amount of interacting components. The present systematic review aims to describe and assess 'best standards' for fidelity assessment for a better articulation of their typology, specifically in the context of multi-component interventions for people with severe mental health difficulties. The content and psychometric properties of existing measures are described, common practices identified, and neglected areas of assessment highlighted, so as to inform future measurement development and a move towards a more systematic approach to fidelity validation.

Introduction

Implementation science argues that, for interventions to be successfully –and sustainably– implemented, they need to be described in sufficient detail to establish the presence and strength of their specific components (Proctor et al., 2009; Forgatch & DeGarmo, 2011). Programme fidelity –or the extent to which an intervention is being delivered as intended in its implementation protocol– is one useful approach to understanding an intervention’s critical components (Bond et al., 2000a; Cabaniss, Wainberg, & Oquendo 2015; Lloyd-Evans et al., 2016; McGrew & Griss, 2005; Schoenwald & Garland, 2013; Teague, Mueser, & Rapp, 2012; Wilson et al., 2009). Further, programme fidelity is associated to treatment outcomes and has a direct impact on an intervention’s internal, external, and construct validity (Borelli, 2011; Gearing et al., 2011; Santacroce, Maccarelli, & Grey, 2004). Without fidelity checks, treatment outcomes may be over- or underestimated (potentially leading to the rejection of effective treatments or the acceptance of ineffective ones) or, conversely, outcomes may be attributed to otherwise irrelevant aspects of an intervention (Moncher & Prinz, 1991).

Fidelity measurement can be a particular challenge for psychosocial interventions for people with severe mental health difficulties. This is primarily because of the large number of interacting contextual, organizational, and service-level components (Wheeler et al., 2015). Multi-component mental health interventions are therefore those where different services, treatment approaches, professional agencies, and disciplines are involved and interact with each other to address the diverse range of client needs. Developing measures that capture these many variables can become a problematic and time-consuming task, especially as there are currently no guidelines about their content and degree of specificity (Bond

et al., 2000a, 2011; Teague et al., 2012; Wilson et al., 2009).

Donabedian (1988) suggested a structure-process-outcome framework to fidelity evaluation, arguing that “a good structure (i.e. material resources, human resources, and organizational structure) increases the likelihood of good processes (i.e. client and practitioner care-related activities), which in turn increase the likelihood of good outcomes (i.e. effects of care on health status)”. This might be more readily identifiable in interventions delivered by a single service provider or practitioner (e.g. CBT for psychosis; Fowler, Garety, & Kuipers, 1995); however, multi-component interventions (e.g. the Assertive Community Treatment or ACT model; Stein & Test, 1980) can make this task burdensome. The structure-process-outcome framework (Donabedian, 1988) is often cited in fidelity literature; however, it is unclear whether existing measures do, in fact, address these domains or whether any particular domains are given priority over others.

The present chapter presents a systematic review of programme fidelity measures of multi-component mental health interventions for people with severe mental health difficulties. The first part of the chapter explores the concept of programme fidelity and the complexity behind assessing it in multi-component interventions. The second part of the chapter presents the aims and method used, followed by a description of the identified measures alongside their characteristics and common features. In the third and final section, future measure developments will be considered in the light of the findings, alongside some clinical and research implications of the review.

Fidelity measurement

A growing problem in implementation literature, is that the term ‘fidelity’ can be seen to be used interchangeably with other related –yet distinct– treatment

evaluation terms, such as ‘adherence’ and ‘integrity’ (Carroll et al., 2007; Cross & West, 2011). Faw (2003) suggested that this linguistic problem could be due to a lack of understanding of the many variables involved in a programme’s ‘process evaluation’ (i.e. everything a programme theoretically and practically involves). Similar to programme fidelity, implementation evaluations aim to assess the degree to which a treatment is implemented as planned (Gresham et al., 1993). When assessed as a dichotomous variable (i.e. presence or absence of certain treatment activities) implementation evaluations are often called ‘treatment fidelity’ evaluations (Corbett, Thompson, White, & Taylor, 1991); however, when considered as multi-dimensional, they are sometimes referred to as ‘fidelity analysis’ or ‘treatment integrity’ evaluations (e.g. Perepletchikova, Treat, & Kazdin, 2007; Sanetti & Kratochwill, 2018).

Multidimensional implementation evaluations can include several additional variables such as, once again, ‘fidelity’ and ‘adherence’. The confusion lies, however, in the language used to define them. ‘Fidelity’ (which will be referred to as ‘programme fidelity’ from this point forward) is often defined as a service’s adherence—in the sense of observance—to an implementation protocol (Pentz et al., 1990). Adherence on the other hand, can be thought of as a narrower sub-category of a fidelity assessment, which focuses on whether the therapeutic techniques, skills, and strategies (e.g. behavioural activation, cognitive restructuring, etc.) are deployed in the experimental group and not in the control group, as set out in a treatment manual (Pentz et al., 1990). Other terms such as ‘dose’ or ‘exposure’ (i.e. frequency and duration of intervention), ‘participant responsiveness’ (i.e. engagement of participants), ‘coverage’ (i.e. whether people for who the intervention was designed for actually receive it), and ‘differentiation’ (i.e. the presence of an intervention’s

essential components) are also considered part of ‘fidelity analyses’ (Carroll et al., 2007; Dusenbury, Brannigan, Falco, & Hansen, 2003; Mihalic, 2004).

This confusion in terminology seems to have been further intensified by a lack of clarification of the level of study. For example, there is often a lack of clarity as to whether these terms pertain at an interventionist level (i.e. delivery agent or individual service provider) or a process level (i.e. everything involved in a treatment programme, including structural, organizational, theoretical and practical components) (Cross & West, 2011). Psychotherapy research, which usually focuses on single treatment sessions and therapist-patient dynamics (Faw, 2003), is an example of interventionist-level evaluation (Cross & West, 2011). In such cases the categorical distinctions might be minimal as the structural (e.g. time length of sessions, treatment duration), organizational (i.e. single provider, single client) and functional (e.g. techniques, skills, and competence) aspects of treatment protocols rely primarily on the agent delivering them, in this case the individual therapist (Cross & West, 2011). In contrast, multi-component interventions require a greater focus on process (Cross & West, 2011). Programme fidelity therefore refers to the observance of a model’s principles, standards, and procedures at all levels, including structural and organizational components (e.g. setting, caseload, staffing, training, pathways, services provided, etc.), and functional processes of assessment and treatment provision (e.g. therapist adherence, therapist competence, assessment areas, etc.) (Bond et al., 2001, 2012, 2016, 2017; Bruns et al., 2004, 2005, 2008; Faw, 2003; Johnson, 2011; Kernan, 2014; McHugo et al., 2007; Pullmann et al., 2013; Rollins et al., 2010).

Programme fidelity in multi-component interventions

Multi-component mental health interventions for people with severe mental

health difficulties have been the focus of attention in the last 20 years and there have recently been calls for increased research and developments in the field (Patel et al., 2018). Although intervention research has led to the development of numerous interventions that have proven effective under rigorous methodologies and highly controlled scenarios, these are not always successfully transported to real-life contexts (see ‘science-to-service gap’ in Drake & Essock, 2009). Translating interventions from controlled trials to community settings requires making multiple adaptations and considering a series of variables that might not otherwise lend themselves for replication during trials (Proctor et al., 2009). Given their micro-analytic nature, programme fidelity evaluations might help bridge this gap.

Programme fidelity assessment is now recognised as an important quality assurance procedure and potentially as a marker of successful implementations (Proctor et al., 2009; Salyers, & Tsembris, 2007; Schoenwald, & Garland, 2013; Torrey, Finnerty, Evans, & Wyzik, 2003; Waghorn, 2009). This might be because programme fidelity may be useful in teasing apart which components of an intervention lead to certain desired outcomes (Bloch, Saed, Rivard, & Rausch, 2006; Torrey, Bond, McHugo & Swain, 2012). Programme fidelity can thus help to (1) enhance an intervention’s efficacy; (2) address areas of weakness and improve adherence; (3) measure service performance over time; (4) replicate interventions through clearer manuals, training, and supervision methods; and (5) disseminate these active ingredients to inform other evidence-based practices (Aarons, Hurlburt, & Horowitz, 2010; Becker et al., 2001; Bond et al., 1997, 2011; Rollins et al., 2010; Teague et al., 1998).

Programme fidelity has also been associated to the effectiveness of an intervention (Faw, 2003). Evidence suggests that higher fidelity scores are strongly

correlated with better client outcomes (Bond & Salyers, 2004; Bond, Smith, Tanzman, Drake, & Tremblay, 2011; McGrew, Bond, Dietzen, & Salyers, 1994; Teague et al., 2012). For example, in the context of multi-component interventions for severe mental health difficulties, clients from high programme fidelity ACT services have been found to have higher rates of remission from substance misuse disorders and a reduction in alcohol and drug use compared to those from low fidelity teams (McHugo, Drake, Teague, & Xie, 1999). Also, high fidelity ACT programmes have been associated with a reduced number of hospital days (Latimer, 2002) and lower perceived coercion (McGrew, Wilson & Bond, 2002) than lower fidelity programmes. This pattern has also been found in the context of supported employment, where lower fidelity scores to the Individual Placement and Support (IPS) model (Becker & Drake, 1993) have been found to correlate with poorer employment outcomes (Bond, Drake, & Becker, 2008). Further, in the context of child and adolescent mental health, high fidelity to the ‘wraparound’ model (Burchard, Burns, & Burchard, 2002) has been found to correlate with child behavioural strengths and a reduced number of unmet needs (Bruns, Suter, Force, & Burchard, 2005). Given their association with treatment outcomes, programme fidelity checks may therefore be relevant to many stakeholders such as (a) funding bodies (who need reassurance that their investment is reaching the expected intervention), (b) service managers (to know how to better distribute resources), (c) clinicians (to identify strengths and areas for improvement), (d) clients (who expect to see the desired outcomes), and (e) even larger institutions (in order to establish accreditation and licensing criteria) (Bond & Drake, 2017; Essock et al., 2015; Teague et al., 2012).

Notwithstanding these benefits, programme fidelity assessment is uncommon

in implementation studies of multi-component mental health interventions (Nugter et al., 2016; Perplechikova et al. 2007; Petrakis et al., 2011). Teague and colleagues (2012) attributed this scarcity to a lack of understanding of the topic and an underlying assumption that describing psychosocial interventions in terms of active components may miss subtler process variables. Others, such as Schoenwald and colleagues (2011), argue this is due to economic reasons, given that current programme fidelity measures are not efficient enough and can be quite resource-intensive.

Echoing Schoenwald's and colleagues (2011) reflections, programme fidelity measures must strike a balance between their effectiveness (i.e. how reliably and validly they capture the "essence" of an intervention) and their efficiency or feasibility in routine practice (Teague et al., 2012). Some measures focus on more easily-measurable structural aspects of service delivery such as caseload, number of sessions, location, staffing, among others, but sacrificing more abstract process variables (e.g. ethos, model of care, etc.) and therefore potentially validity (e.g. Nugter et al., 2016). Other measures might focus on the nature of human interactions, perhaps by using video-recorded sessions or observations of sessions, yet sacrificing efficiency or potentially the reliability of findings (Drake & Deegan, 2008).

Considering these challenges, implementation research suggest that best practice in programme fidelity assessment would involve (1) an evidence-based, comprehensive, and multimodal approach to assessment, (2) with clearly and objectively operationalised components stemming from a coherent and comprehensive theory of change, and (3) easily-available data (e.g. service records, operational policies, administrative checklists, etc.) from the relevant stakeholders

(Essock et al., 2015; Schoenwald, & Garland, 2013; Waghorn, 2009). In the context of severe mental health difficulties, Bond and colleagues (2000a) developed a toolkit to support the development of programme fidelity measures for psychiatric rehabilitation interventions (e.g. ACT). The toolkit included a thorough review and analysis of available measures with useful recommendations on how to develop effective and efficient scales. However, many of the measures reviewed were concerned with what we now might call ‘adherence’ (i.e. interventionist level) rather than focusing on ‘programme fidelity’ itself. Further, there have been no additional reviews of available measures for multi-component interventions for severe mental health difficulties since the year 2000. It is possible, given the publication times of the literature, that the aforementioned confusion of terms has not yet been taken into consideration, thus leading to disagreements regarding what programme fidelity measures ought to focus on.

Aims

The present systematic review of programme fidelity measures of multi-component interventions for people with severe mental health difficulties builds on the call for a better articulation of the typology of fidelity measurements (Teague et al., 2012).

The aim of this chapter is to identify the existing measures and describe their characteristics (i.e. interventions assessed, number of items, structure, format, scoring procedures, participants, data sources, time to administer, requirements, development procedures), content (i.e. domains assessed), and psychometric properties (i.e. reliability and validity) in order to identify common practices and gaps in knowledge. A narrative synthesis approach is used to summarize findings. This intention is to bring greater clarity to the concept and definition of fidelity and help future developments of measures and assessment procedures.

Method

Design

A systematic review (PROSPERO registration No. CRD42018108360) was conducted following Cochrane methodology (Higgins & Green, 2011) and recommendations by the COSMIN guideline for systematic reviews of PROMs (Prinsen et al., 2018).

Inclusion criteria

Population. Children and adults with severe mental health difficulties. The term “severe mental health difficulty” is particularly difficult to delineate given its dependence on culturally-bound ideas of health and illness (Patel et al., 2018); it was therefore broadly defined as “any mental, behavioural, or emotional disorder resulting in serious functional life impairment” (Hazelden Foundation, 2008; National Institute of Mental Health, 2019; Patel et al., 2018; Public Health England, 2018). As such, the term included the diagnoses of schizophrenia or schizophrenia-like disorders, bipolar disorder, substance misuse, depression with psychotic features, eating disorders, personality disorders, suicide, self-harm, and conduct disorders, however diagnosed or defined and whether acute or chronic. Cohort and case-control studies from inpatient, community, and outpatient services were included. Likewise, studies of mixed populations were included, if a majority of participants had a severe mental health diagnosis.

Intervention. Multi-component mental health interventions, defined as interventions with multiple interacting components (e.g. services, treatment approaches, agencies, types of professionals and disciplines involved) that include a core mental health component in combination with one or more of the following domains: physical health treatment, employment or educational support, criminal

justice services, or external agencies; for a single or multiple types of severe mental health difficulties in inpatient or outpatient contexts.

Properties. Psychometric properties included –where available– were reliability (i.e. inter-rater reliability, internal consistency, test-retest reliability) and validity (i.e. face validity, content validity, construct validity, and criterion or predictive validity).

Outputs. Service-level (i.e. programme) fidelity measures.

Exclusion criteria

1. Studies with a sole focus on samples with neurodevelopmental or neurocognitive difficulties, and learning disabilities.
2. Studies without evaluation of both the content and psychometric properties of the measure(s) used.
3. Studies focusing on single-component or stand-alone interventions (e.g. dialectical-behaviour therapy with no additional components).
4. Studies focusing on one-to-one interventions (e.g. cognitive-behaviour therapy, psychodynamic psychotherapy).
5. Multi-component interventions without an active mental health component.
6. Measures focusing on interventionist-level fidelity (e.g. therapist adherence or therapist competence).

Search strategy

Six electronic databases (MEDLINE and Epub Ahead of Print, In-Process and Other Non-indexed Citations and Daily; EMBASE; PsycINFO; Cochrane Library; Web of Science; and Health and Psychosocial Instruments) were searched from their creation date to November 2018. After the initial search, reference lists of

reviews, relevant protocols, and forward and backward searching of included studies were screened to identify further articles.

Initial search terms were piloted and refined iteratively with sequential testing to identify false-positive and false-negative results, and ensure that the search captured all relevant keywords. Free and mapped searches, using Medical Subject Heading (MeSH) terms, were conducted. Database terms used included “mental disorders”, “bipolar”, “psychotic disorders”, “eating disorders”, “personality disorder”, “substance related disorders”, and “depressive disorder, major”. Boolean operators were used to construct a search incorporating all search terms when combination searches were not possible. The PEERS checklist (McGowan et al., 2016) was used to enhance the quality of the search strategy. Search outputs were filtered for English full texts and peer-reviewed articles. The final search strategy is on Appendix A.

Study selection

Two reviewers (Mauricio Alvarez and Melissa Lotmore, Trainee Clinical Psychologists) conducted the electronic searches and screened the reference lists of relevant articles. All identified titles and abstracts were downloaded and merged using EndNote X8 and de-duplicated prior to screening. Due to the volume of initial records it was decided that the same reviewers would independently screen a random 10% sample ($n = 623$) of the total records found, only at title level. Since inter-rater reliability between raters was very good (93.1% agreement, Cohen’s kappa = 0.76) it was agreed that the remainder of title screenings would be performed individually. Reviewers met once all titles had been screened to resolve discrepancies and the senior systematic reviewer (Prof. Stephen Pilling) acted as an arbiter where necessary. Since MA and ML were looking into different aspects of treatment

integrity assessment measures (i.e. adherence and fidelity) abstracts and full texts were screened individually against inclusion and exclusion criteria and results were discussed afterwards to discuss any overlap. Any articles which reviewers were unsure of were retained until data extraction, when more information was available (Higgins & Deeks, 2008).

Data extraction

A data extraction template (Appendix B) was designed based on recommendations from the COSMIN guideline (Prinsen et al., 2018). Data was tabulated by both extractors to aid synthesis (Appendix 2). The data to be extracted (where available) was the following: (1) country, (2) instrument name, (3) availability of measure manual, (4) intervention assessed, (5) disorder treated in the intervention, (6) age group, (7) measure description, (8) definition of fidelity used, (9) domains covered, (10) number of items, (11) scoring procedure, (12) participants, (13) data sources, (14) type of rater (i.e. measure completer), (15) time to administer, (16) training requirements, (17) reliability, (18) validity, and (19) development methods.

Risk of bias assessment

Study quality was not formally assessed due to lack of appropriate standardised tools to assess bias in fidelity measure validation studies. However, the quality of measure properties was established using recommendations from the COSMIN guideline (Mokkink et al., 2018). Studies deemed to be of poor methodological quality would be discussed with SP.

Data synthesis

Due to the heterogeneity of studies, the content, characteristics, and psychometric properties of the fidelity measures found were summarized using a

narrative synthesis approach (Popay et al., 2006). Additionally, a preliminary thematic analysis (Braun & Clarke, 2006) was used to identify domain similarities across measures. This analysis used semantic (i.e. explicit) themes, an inductive (i.e. data-driven) approach, and an essentialist (i.e. unidirectional interpretation of meaning) epistemology. Themes and domains were established by the main author based on the description of each measure's dimensions and discussed with the supervisor for face and content validity.

Results

After duplicates were removed, 6235 records were identified. Thirty-one records were included in the analysis (Appendix C).

Study characteristics

All of the studies included described programme fidelity measures for multi-component interventions for severe mental health difficulties. Seven papers (22.6%) were focused on child and adolescent mental health services (CAMHS) and twenty-four (77.4%) related to adult mental health services.

Fidelity measures identified

Twelve measures were extracted from the included papers (Table 1). Ten (83.3%) of these measures were developed in the United States, while one (8.3%) was developed in Netherlands and one (8.3%) in the United Kingdom. These studies took place in the settings of psychiatric rehabilitation (n=7, 58.3%), supported employment (n=3, 25%), and CAMHS (n=2, 16.7%). Identified measures assessed programme fidelity of the following multi-component interventions: Therapeutic community (n=1, 8.3%), Crisis Resolution Teams (n=1, 8.3%), Assertive Community Treatment (n=3, 25%), Integrated Treatment for Dual Disorders (n=2, 16.7%), Individual Placement and Support/Supported employment (n=3, 25%), and

Table 1. Programme fidelity measures.

Measure	References	Setting	Intervention	Country
1. Adolescent Treatment Program (ATP) Environment Scale	Faw, 2003	Child and adolescent mental health	Therapeutic Community	USA
2. Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services (CIMHRRS)	Johnson, 2011	Psychiatric rehabilitation	No specific intervention. Focus on overarching rehabilitation principles.	USA
3. Crisis Resolution Team Fidelity Scale (CORE CRT)	Lloyd-Evans et al., 2016	Psychiatric rehabilitation	Crisis Resolution Teams	United Kingdom
4. Dartmouth Assertive Community Treatment Fidelity Scale (DACTS)	Johnsen et al., 1999 Kidd et al., 2010 McGrew et al., 2013 McHugo et al., 2007 Rollins et al., 2010 Salyers et al., 2003 Teague et al., 1998 Winter & Calsyn, 2000	Psychiatric rehabilitation	Assertive Community Treatment	USA
5. Dual-Disorder Treatment Fidelity Scale (DDT)	Wilson et al., 2009	Psychiatric rehabilitation	Integrated Treatment for Dual Disorders	USA
6. Flexible Assertive Community Treatment Scale (FACTS)	Nugter et al., 2016	Psychiatric rehabilitation	Flexible Assertive Community Treatment	Netherlands
7. Index of Fidelity of Assertive Community Treatment (IFACT)	McGrew et al., 1994	Psychiatric rehabilitation	Assertive Community Treatment	USA

Table 1 (Continued). Fidelity measures.

Measure	References	Setting	Intervention	Country
8. Individual Placement and Support Fidelity Scale (IPS)	Bond et al., 1997 (IPS-15) Bond et al., 2001 (IPS-15) Becker et al., 2001 Bond et al., 2011 McGrew et al., 2005	Supported employment	Individual Placement and Support	USA
9. Individual Placement and Support – 25 item scale (IPS-25)	Bond et al., 2012 Bond et al., 2016	Supported employment	Individual Placement and Support	USA
10. Integrated Dual Disorders Treatment Fidelity Tool (IDDT)	Harrison et al., 2017 McHugo et al., 2007	Psychiatric rehabilitation	Integrated Treatment for Dual Diagnosis	USA
11. Quality of Supported Employment Implementation Scale (QSEIS)	Bond et al., 2000b Bond et al., 2002 McGrew et al., 2005	Supported employment	Supported Employment	USA
12. Wraparound Fidelity Index (WFI)	Bruns et al., 2004 Bruns et al., 2005 Bruns et al., 2008 Effland et al., 2011 Kernan, 2014 Pullmann et al., 2013	Child and adolescent mental health	Wraparound	USA

Wraparound (n=1, 8.3%). For a summary of findings please see appendices D-G.

1. Adolescent Treatment Program (ATP) Environment Scale (Faw, 2003). The ATP measures the extent to which inpatient services providing the Adolescent Treatment Program for substance misuse in adolescents adhere to the principles of a therapeutic community. It consists of 21 dichotomous items assessing the service's adherence to 6 treatment principles, namely (1) peers as gatekeepers, (2) mutual help, (3) enhancement of community belonging, (4) contact with outside community, (5) community/clinical management, and (6) 'level system' (i.e. privileges and responsibilities based on behavioural contingencies). The ATP Environment Scale is completed weekly by clients for the duration of the intervention.

2. Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services (CIMHRRS) (Johnson, 2010). The CIMHRRS characterizes the structural, functional, and organizational process differences of settings addressing severe mental health difficulties (Johnson, 2011). It consists of 52 items covering 8 principles: (1) program mission, (2) program demographics and composition, (3) organizational boundaries, (4) program functioning, (5) treatment team structure and process, (6) assessment process, (7) treatment planning, and (8) treatment provision. Two trained raters score each item on a 5-point behaviourally-anchored Likert scale. The CIMHRRS can take up to 16 hours to complete per site.

3. CORE Crisis Resolution Team Fidelity Scale (CRT) (Lloyd-Evans et al., 2016). The CORE CRT measures the implementation of the crisis and recovery team (CRT) model. It consists of 32 items covering 4 dimensions: (1) referrals and access, (2) content and delivery of care, (3) staffing and team procedures, and (4) timing and location of care. Three trained raters score each item on a 5-point

behaviourally-anchored Likert scale. The CORE CRT is a complex measure that can take about a full day to complete per site.

4. Dartmouth Assertive Community Treatment Fidelity Scale (DACTS) (Teague et al., 1998). The DACTS is the most widely used measure for the ACT model. The DACTS is an adaptation and formalization of a more experimental measure to evaluate the Programme in Assertive Community Treatment (PACT) (Teague, Drake, & Ackerson, 1995). The DACTS originally consisted of 26 items covering three broad dimensions: (1) human resources, (2) organizational boundaries and/or structure, and (3) nature or delivery of services. A later expansion included two additional items focusing on staff size (HR11) and the role of consumers (S10); however, since they were included after the main validation study, they are only included for ACT service fidelity evaluations (McGrew, White, Stull, & Wright-Berryman, 2013). Each item on the DACTS is scored on a 5-point behaviourally-anchored Likert scale and a final score is arrived after averaging all items. The DACTS consists of a thorough, day-long assessment, carried out by two trained raters. A study by McGrew and colleagues (2013) found that the DACTS could also be reliably scored when using only a set of 9 self-report survey tables.

Recent studies have gradually replaced the DACTs with the Tool for the Measurement of ACT (TMACT) developed by Monroe-DeVita and colleagues (2011). The TMACT is a more comprehensive measure (47 items) that covers (1) operations and structure, (2) the core team, (3) the specialist team, (4) core practices, (5) evidence-based practices, and (6) person-centred planning and practices. As such, TMACT is considered of a much higher standard than the DACTS; unfortunately, since our search did not come across any validation studies with psychometric properties it was not included in the current review.

5. Dual-Disorder Treatment Fidelity Scale (DDT) (Mueser et al., 2003).

The DDT assesses the robustness of system-, programme-, and clinician-level implementation of dual-disorder treatment. This unidimensional measure consists 20 items scored by two trained raters (40-hour training) on a 5-point behaviourally-anchored Likert-type scale.

6. Flexible Assertive Community Treatment Scale (FACTS) (Nugter et

al., 2016). The FACTS is an adaptation of DACTS to assess the level of implementation of the Flexible ACT model. The FACTS consists of 60 items covering 7 dimensions: (1) team structure, (2) team process, (3) diagnostics & interventions, (4) organization of services, (5) level of social services, (6) use of routine outcome measures, and (7) level of professionalization. All items are scored by two trained raters on a 5-point behaviourally-anchored Likert-type scale and its completion takes an average of 45 minutes per site.

7. Index of Fidelity of Assertive Community Treatment (IFACT)

(McGrew et al., 1994). The IFACT was the first measure developed to assess fidelity to the ACT model and from where later adaptations –such as the PACT, DACTS, FACTS, and TMACT– emerged. It consists of 17 items covering three dimensions: (1) staffing, (2) organization, (3) service. Each item is scored using a minimum thresholds approach (i.e. continuous, zero-to-one scale). The only paper assessing IFACT, however, used pre-existing datasets of ACT services and therefore no details are available regarding the time requirements, rater specifications or training needs.

8. Individual Placement and Support (IPS) scale (Bond et al., 1997).

The IPS assesses the degree to which specific programmes meet the standards of the IPS model of supported employment and identify the critical ingredients of the IPS. The

original version, the IPS-15, consists of 15 items covering 3 dimensions: (1) staffing, (2) organization, and (3) services. Each item is scored by two trained raters on a 5-point behaviourally-anchored Likert-type scale. IPS-15 completion can take up to 60 minutes per site.

The IPS-15 was later consolidated –and renamed for proceeding studies– as the Individual Placement and Support Fidelity Scale (IPS) (Bond et al., 2011; McHugo et al., 2007). This version is virtually identical to its preceding version but includes new data sources. As a result, the IPS scale takes up to 90-120 minutes to complete, which is usually done over a day-long site visit. Follow-up studies found that doing telephone interviews (McGrew et al., 2005) or using self-reported surveys to gather information (Becker et al., 2001) was equally reliable as the face-to-face approach and could reduce the time requirements of the measure.

9. Individual Placement and Support (IPS-25) scale (Becker et al., 2008). The IPS-25 is an expansion of the IPS scale (Bond et al., 1997) to improve the instrument’s internal consistency. This unidimensional measure consists of 25 items, which means it can take up to a day and a half to complete per site. Like its previous version, the IPS-25 is scored by two trained raters on a 5-point behaviourally-anchored Likert scale.

10. Integrated Dual Disorders Treatment Fidelity Tool (IDDT) (McHugo et al., 2007). The IDDT assesses the level of fidelity to the principles and procedures of the fully integrated dual disorder treatment model (McHugo et al., 2007). It consists of 26 items covering 2 broad dimensions: (1) organization and (2) treatment; however, it was designed to be used in conjunction with the General Organizational Index, a 12-item measure that assesses general agency practices that support evidence-based practices (Drake et al., 2001). Following the scoring and cut-

off procedures of other fidelity tools, the IDDT tool is scored by two-to-three trained raters on a 5-point behaviourally-anchored Likert-type scale. Completion is usually done over a day-long site visit.

11. Quality of Supported Employment Implementation Scale (QSEIS)

(Bond et al., 2000b). The QSEIS is a checklist used to determine the degree of attainment of practice standards for supported employment in the context of therapeutic communities for clients with dual disorders (Bond et al., 2000b). It was originally developed by Vogler (1998) as part of a doctoral dissertation project and formalised by Bond and colleagues (2000b) into a 33-item measure covering the same 3 dimensions of the IPS scale: (1) vocational staffing, (2) organization, and (3) services. The QSEIS is scored by two trained raters on a 5-point behaviourally-anchored Likert-type scale.

12. Wraparound Fidelity Inventory (WFI) (Bruns et al., 2004). The WFI determines the level of adherence of a service to core principles of the ‘wraparound’ model for adolescents with lived experience of severe mental health difficulties during their transition to adulthood (Walker & Bruns, 2013). The WFI consists of three measure formats covering information from care facilitators (CF), caregivers (CG) and youth (Y). The CF and CG forms are composed of 44 items and 32 items for the youth form, and they cover the following 10 dimensions or principles: (1) family voice and choice, (2) youth and family team, (3) community-based services, (4) cultural competence, (5) individualized, strength-based services, (6) natural supports, (7) continuation of care, (8) collaboration, (9) flexible resources, and (10) outcome-based services. Each item is scored by a single trained rater on a 3-point Likert-type scale (i.e. no, somewhat, yes). Although authors originally recommend that two raters should be involved in order to improve the reliability of scores,

Pullman and colleagues (2013) reported this only led to small improvements. Bruns and colleagues (2004) reported an average administration time of 56.9 minutes but it has been reported that the latest version (WFI-4) can take between 15 and 40 minutes to complete per site (Kernan, 2014).

Measure properties

Structure and design. There was a considerable variability in the number of items included in each measure (range=15-60, mean=31.67, median=27). All measures had two or more sources of information such as on-site or telephone interviews, observations of meetings and interventions, chart reviews or documentation. Eleven measures (91.7%) were intended to be scored by 2 or more raters, and only one (8.3%) was self-rated (Appendix E). In terms of scoring, 10 measures (83.3%) used behaviourally-anchored Likert-type scales, one (8.3%) used a dichotomous (i.e. present or absent) scale, and one (8.3%) used a minimum thresholds approach (i.e. scoring one when the item meets or exceeds expert criterion). Only 5 measures (41.7%) included scoring cut-offs (see Appendix D for a more detailed account).

Psychometric properties. There was great variability in the amount and level of detail of psychometric properties reported. In relation to reliability, most measures (n=9, 75%) reported internal consistency and inter-rater reliability (n=8, 66.7%); however, only two measures reported test-retest reliability (n=2, 16.7%). Overall, one measure (8.3%) had excellent reliability (i.e. CORE CRT); 5 measures (41.7%) had very good reliability (i.e. CIMHRRS, DACTS, DDT, FACTS, and QSEIS); 5 measures (41.7%) had good reliability (i.e. IFACT, IPS, IPS-25, IDDT, and WFI); and one measure had adequate reliability (i.e. ATP). In relation to validity, most measures (n=7; 58.3%) reported face/construct and criterion validity

and only some (n=5, 41.7%) reported content validity. Overall, 9 measures (75%) had good validity (i.e. CIMHRRS, CORE CRT, DACTS, FACTS, IFACT, IPS, IDDT, QSEIS, and WFI); two measures (16.7%) had acceptable validity (i.e. IPS, and DDT); and one measure (8.3%) had limited validity (i.e. ATP) (see Appendix F for a detailed account).

Thematic analysis: Assessment domains

Three broad domains were identified across measures: (1) process (i.e. policies, ethos, and community orientation), (2) organizational structure (i.e. operations, staff, and team management), and (3) delivery of care (i.e. services, interventions, and monitoring) (Table 2). When analysed on a domain level there was variation across measures, with 9 (75%) measures focusing on the organizational structure of services, 6 (50%) measures on the delivery of care, and three (35%) measures focusing on process features. However, when analysed on an item-level a more even distribution was observed, with all measures (100%) including one or multiple items for each domain.

After sorting all items across measures, 50 different features (i.e. themes) could be identified across domains (Tables 3-5). All measures had a different configuration of items. Overall, the CORE CRT and the FACTS appeared to be the most comprehensive measures, covering 29 of the 50 features identified (58%), whereas the IPS and the DDT measures were the least comprehensive ones, with only 10 (20%) and 11 (22%) features covered respectively.

Process. Process items focused on policies (i.e. mission, availability, access, referrals, admission criteria, and safety), programme ethos (i.e. model of care, philosophy, penetration, choice, and cultural competence), and community orientation of the programme (i.e. service location, home-visit arrangements,

Table 2. Common domains across measures.

Measure	Process			Organizational structure			Delivery of care		
	Policies	Ethos	Community orientation	Operations	Staff	Team	Services	Interventions	Monitoring
1. ATP	-	X	X	-	-	-	X	-	-
2. CIMHRRS	X	-	-	X	-	-	X	-	X
3. CORE CRT	-	-	-	X	X	-	X	X	-
4. DACTS	-	-	-	X	X	X	-	-	-
5. DDT*	-	-	-	-	-	-	-	-	-
6. FACTS	-	-	-	-	X	X	X	-	X
7. IFACT	-	-	-	X	X	X	-	-	-
8. IPSFS	-	-	-	X	X	X	-	-	-
9. IPS-25*	-	-	-	-	-	-	-	-	-
10. IDDT	-	-	-	X	-	-	X	-	-
11. QSEIS	-	-	-	X	X	X	-	-	-
12. WFI	-	X	X	-	-	-	X	X	X
Frequency	1	2	2	7	6	5	6	2	3
%	8.3	16.7	16.7	58.3	50	41.6	50	16.7	25
Weighted total		3			9			6	
%		35			75			50	

Notes: Numbers represent the number of domains included for each feature in each measure. ATP = Adolescent Treatment Program Environment Scale; CIMHRRS = Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services; CORE CRT = Crisis Resolution Team Fidelity Scale; DACTS = Dartmouth Assertive Community Treatment Scale; DDT = Dual Disorder Treatment Fidelity Scale; FACTS = Flexible Assertive Community Treatment Scale; IFACT = Index of Fidelity of Assertive Community Treatment; IPS = Individual Placement and Support Fidelity Scale; IPS-25 = Individual Placement and Support – 25 item scale; IDDT = Integrated Dual Disorders Treatment Fidelity Tool; QSEIS = Quality of Supported Employment Implementation Scale; WFI = Wraparound Fidelity Index

* Unidimensional measures.

Table 3. Process features by measure (item level).

	CIMHRRS	CORE CRT	DACTS	DDT	FACTS	IFACT	IPS	IPS-25	IDDT	QSEIS	WFI
Policies											
1. Mission	X	X	-	-	-	X	X	X	-	X	X
2. Availability	-	X	X	X	X	X	-	X	X	-	-
3. Access	-	X	X	-	X	-	X	X	-	-	-
4. Referral	-	X	-	-	-	-	-	-	-	X	-
5. Admission criteria	X	-	X	-	X	-	-	-	X	-	-
6. Safety	-	X	-	-	-	-	-	-	-	-	-
Ethos											
7. Model	X	-	X	-	-	-	X	X	-	X	-
8. Philosophy	X	X	-	X	X	-	-	X	X	X	X
9. Penetration	-	-	-	-	-	-	-	-	X	-	-
10. Choice	X	X	X	-	-	-	X	-	X	-	X
11. Cultural competence	-	X	-	-	-	-	-	-	-	-	X
Community orientation											
12. Location	-	X	X	-	-	-	X	X	-	X	-
13. Home visits	-	X	-	-	-	X	-	-	-	-	-
14. Reinsertion	-	-	-	-	-	-	-	-	-	-	X
15. System involvement	-	X	X	-	X	-	-	X	-	X	X
16. System supports	-	-	-	-	X	-	-	-	-	-	X
17. Carers' support	-	X	-	-	-	-	-	-	-	-	-
Total	5	12	7	2	6	3	5	7	5	6	7
%	17.6	29.4	80	41.2	11.8	35.6	17.6	29.4	41.2	29.4	41.2

Notes: ATP = Adolescent Treatment Program Environment Scale; CIMHRRS = Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services; CORE CRT = Crisis Resolution Team Fidelity Scale; DACTS = Dartmouth Assertive Community Treatment Scale; DDT = Dual Disorder Treatment Fidelity Scale; FACTS = Flexible Assertive Community Treatment Scale; IFACT = Index of Fidelity of Assertive Community Treatment; IPS = Individual Placement and Support Fidelity Scale; IPS-25 = Individual Placement and Support – 25 item scale; IDDT = Integrated Dual Disorders Treatment Fidelity Tool; QSEIS = Quality of Supported Employment Implementation Scale; WFI = Wraparound Fidelity Index.

Table 4. Organizational structure features by measure (item level)

	CIMHRRS	CORE CRT	DACTS	DDT	FACTS	IFACT	IPS	IPS-25	IDDT	QSEIS	WFI
Operations											
18. Caseload size	X	-	X	-	X	-	X	X	-	X	-
19. Client-staff ratio	X	-	X	X	X	X	-	-	-	X	-
20. Multiagency integration	X	-	-	X	X	-	X	X	-	X	-
21. Information sharing	-	X	-	-	X	X	-	X	-	X	-
22. Client turnover	X	X	X	-	-	-	-	-	-	-	-
23. Waiting list	-	-	-	-	X	-	-	-	-	X	-
24. Treatment length	-	X	-	-	-	X	-	-	-	X	-
25. Clinical contacts	-	X	X	-	X	-	-	X	-	X	-
Staff											
26. Population served	X	-	-	-	-	-	-	-	-	-	-
27. Staffing & turnover	X	X	X	-	X	X	-	-	-	-	-
28. Roles	X	X	X	-	X	X	-	-	X	X	-
29. Training	X	-	-	-	X	-	-	-	X	-	-
30. Supervision	-	-	-	-	-	-	-	X	X	-	-
Team management											
31. Team approach	X	X	X	-	X	X	X	X	X	X	X
32. Leadership	-	-	X	-	X	X	-	-	-	-	-
33. Team Meetings	-	-	X	-	X	X	-	-	-	X	-
Total	9	7	9	2	12	8	3	6	4	10	1
%	56.3	43.8	56.3	12.5	75	50	18.8	37.5	25	62.5	6.3

Notes: ATP = Adolescent Treatment Program Environment Scale; CIMHRRS = Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services; CORE CRT = Crisis Resolution Team Fidelity Scale; DACTS = Dartmouth Assertive Community Treatment Scale; DDT = Dual Disorder Treatment Fidelity Scale; FACTS = Flexible Assertive Community Treatment Scale; IFACT = Index of Fidelity of Assertive Community Treatment; IPS = Individual Placement and Support Fidelity Scale; IPS-25 = Individual Placement and Support – 25 item scale; IDDT = Integrated Dual Disorders Treatment Fidelity Tool; QSEIS = Quality of Supported Employment Implementation Scale; WFI = Wraparound Fidelity Index.

Table 5. Delivery of care features by measure (item level)

	CIMHRRS	CORE CRT	DACTS	DDT	FACTS	IFACT	IPS	IPS-25	IDDT	QSEIS	WFI
Services											
34. Information	-	X	-	-	-	-	-	-	-	X	X
35. Comprehensiveness	X	-	X	X	X	-	-	-	X	-	X
36. Treatment tailoring	X	X	X	X	X	-	-	X	X	X	X
37. Assertive outreach	-	X	X	X	X	-	X	X	X	X	X
38. Risk assessment	X	X	-	-	-	-	-	-	-	-	X
39. Crisis services	X	X	X	X	X	-	-	-	-	-	X
40. Discharge	X	-	X	-	X	-	-	-	-	-	X
41. Aftercare	-	X	-	-	-	-	-	-	-	-	X
Interventions											
42. Mental health	X	X	X	X	X	-	-	-	X	-	-
43. Physical health	-	X	-	-	X	-	-	-	X	-	-
44. Medication	-	X	-	X	X	-	-	-	X	-	-
45. Social support	-	X	-	-	-	-	-	X	-	X	-
46. Vocational support	-	-	-	-	X	X	-	X	-	X	X
Monitoring											
47. Assessment	X	-	-	X	-	X	X	X	-	-	-
48. Progress reviewing	X	-	-	-	-	X	-	-	X	X	X
49. Outcome monitoring	X	-	-	-	X	-	-	-	X	-	X
50. Feedback	-	-	-	-	X	-	-	-	X	X	X
Total	9	10	6	7	11	3	2	5	9	7	12
%	52.9	58.8	35.3	41.2	64.7	17.6	11.8	29.4	52.9	41.2	70.6

Notes: ATP = Adolescent Treatment Program Environment Scale; CIMHRRS = Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services; CORE CRT = Crisis Resolution Team Fidelity Scale; DACTS = Dartmouth Assertive Community Treatment Scale; DDT = Dual Disorder Treatment Fidelity Scale; FACTS = Flexible Assertive Community Treatment Scale; IFACT = Index of Fidelity of Assertive Community Treatment; IPS = Individual Placement and Support Fidelity Scale; IPS-25 = Individual Placement and Support – 25 item scale; IDDT = Integrated Dual Disorders Treatment Fidelity Tool; QSEIS = Quality of Supported Employment Implementation Scale; WFI = Wraparound Fidelity Index.

service-user reinsertion, system involvement, formal and informal supports from the system, and carers' support). The most common process features were: (1) philosophy (n=8, 73%), (2) mission (n=7, 63.6%), (3) availability (n=7, 63.6%), (4) system involvement (n=7, 63.6%), and (5) service-user choice (n=6, 54.5%). The least common process features were: (1) cultural competence (n=2, 18.2%), (2) home visit arrangements (n=2, 18.2%), and (3) community reinsertion (n=2, 18.2%). The most comprehensive measure on a process level was the CORE CRT, covering 12 of the 17 features (80%). The least comprehensive measures were the DDT, which only covered two (11.8%) process features (i.e. availability and philosophy) and the IFACT, which included only three (17.6%) (Table 3).

Organizational structure. Structural items focused on operations (i.e. population served, caseload size, service-user-to-staff ratio, integration of services, information sharing and communication, service-user turnover, waiting list, treatment length, and number of clinical contacts); staff (i.e. staffing and turnover, staff roles, training and supervision); and team management (i.e. team approach, leadership, and team meetings). The most common organizational structure features addressed were: (1) team approach (n=10, 90.9%), (2) staff roles (n=7, 63.6%), 3) caseload size (n=6, 54.5%), 4) client-to-staff ratio (n=6, 54.5%), and 5) multi-agency integration (n=6, 54.5%). The least common organizational structure features addressed were: (1) population served (n=1, 9.1%), (2) waiting list (n=2, 18.2%), and (3) supervision (n=2, 18.2%). The most comprehensive measure on an organizational level was the FACTS, with 12 of the 16 features covered (75%), followed by the QSEIS (10 features, 62.5%), the CMHRRS and the DACTS (both covering 9 features, 53.6%). The least comprehensive measures were the WFI, with only one feature included (6.3%), followed

by the DDT (two features, 12.5%) and the IPS (three features, 18.8%) (Table 4).

Delivery of care. Delivery items focused on services (i.e. providing information to service users, comprehensiveness of services, individualised treatment, assertive outreach, risk assessment, crisis services, discharge, and aftercare); interventions (i.e. mental health interventions, physical health interventions, medication, social support, and vocational support); and monitoring practices (i.e. assessment, progress reviewing, outcome monitoring, and service-user feedback). The most common features of delivery of care were: (1) treatment tailoring (n=10, 90.9%), (2) assertive outreach (n=9, 81.8%), (3) mental health interventions (n=7, 63.6%), (4) comprehensiveness (n=6, 54.5%), and (5) crisis services (n=6, 54.5%). The least common feature of delivery of care was aftercare arrangements, with only two measures (18.2%) addressing them. The most comprehensive measure in relation to the delivery of care was the WFI, with 12 of the 17 features covered (70.6%), followed by the FACTS (11 features, 64.7%), and the CORE CRT (10 features, 58.8%). The least comprehensive measure was the IPS, with only two features covered (11.8%) (Table 5).

Discussion

General findings

As discussed in the introduction, fidelity is a term that has a history of being used interchangeably with other terms such as adherence and integrity. This is why perhaps the first thing to discuss is the substantial amount of hits resulting from the systematic search (n = 11247). The purpose of having an extensive search strategy was to make the search as narrow as possible to exclude unrelated uses of the term (e.g. relational fidelity, image fidelity) while at the same time broad enough to include measures where the terms ‘fidelity’, ‘adherence’, and ‘integrity’ could be being used interchangeably. In

spite of the confusion of terms however, it seemed that, at least on a measurement level, there was a certain consistency in the definitions and methods of programme fidelity assessment (see Appendix G for a detailed account of definitions).

The scarcity of studies focusing on fidelity measurement for interventions in the context of CAMHS also came into consideration. This is an interesting finding given the growing interest of multi-component interventions for young people (e.g. multisystemic therapy for juvenile offenders, Henggeler et al., 1986). However, this disparity could be explained by the difficulty in defining severe mental health difficulties in this population or because interventions in this context might not necessarily be multi-component. Likewise, it is possible that the scope of this review (i.e. exclusion of papers without psychometric properties) could have led to missing relevant measures and not met the purpose of the study.

The large percentage of measures developed in the United States was also an interesting finding. It is unclear whether this reflects a growing interest for multi-component interventions for severe mental health difficulties in other countries or a greater importance given to fidelity measurement in the United States. However, it is possible that these assessments have gained more attention in United States, given the emphasis in the US health care system on protocols for service provision (see Bond & Drake, 2017) and the use of programme fidelity for service accreditation.

It might be important to mention, that although the search did identify multiple studies and measures related to the supported housing approach for veterans with severe mental health difficulties developed by the United States Department of Housing and Urban Development and the Department of Veteran Affairs Supported Housing (HUD-VASH) initiative, these were not included into the final analysis. Although the HUD-

VASH programme was in fact modelled after ACT, it does not involve an active mental health component. Instead its geared towards (1) finding an apartment, (2) negotiating the lease, and (3) furnishing and moving into said apartment (Rosenheck, Kaspro, Frisman, & Mares, 2003).

Measure properties

The common system was a 5-point behaviourally-anchored Likert scale. However, it is important to consider Bond and colleagues (1997) suggestion that a 4-point scale may offer the highest level of precision possible for rating programme fidelity. This could be one of the reasons why some measures struggle to achieve high levels of reliability. While it might be easier to operationalize and agree on structural and organizational features of a service (e.g. staffing levels, turnover, admission criteria) it might be harder to distinguish between higher and lower scores of process or theory-driven variables (e.g. service principles, system involvement, therapeutic approaches). Nevertheless, the finding that most measures had high levels of reliability underscores the importance of having more than one rater to discuss ratings.

In terms of validity, findings were disappointing. In general, the developers of all measures reported acceptable methods for measure development (e.g. Delphi method, concept mapping, expert panels) which suggest that most measures had reasonable face and content validity. However, few measures reported findings of construct and criterion validity. The lack of construct validity may be because of the scarcity of validated reference measures or proxies to compare, whilst the lack of criterion validity may be explained by a neglect in repeating measurements to establish whether the expected outcomes were associated with fidelity scores. Results from this review emphasise not only the importance of developing more efficient fidelity measures but also of validating

them properly and testing them longitudinally. This would lead to more accurate scoring cut-offs and help further understand the fidelity-outcome relationship (Teague et al., 2012).

As discussed in the introduction, programme fidelity measurement should be efficient and sustainable, requiring as little time and resources from the services being assessed (Essock et al., 2015). This can be challenging whilst at the same time seeking to address as many domains as possible to properly understand the services under question. Findings from this review suggest that measures can greatly vary in terms of the amount of items included and time required to complete them; however, considering the most reliable and valid measures (i.e. CORE CRT, DACTS, FACTS, and QSEIS), it might be possible to develop an appropriate item-to-time ratio for future measures.

The CORE CRT had the highest reliability and validity thus far but requires up to one day to complete 52 items, whereas the longest –and marginally less reliable– FACTS only requires up to 90 minutes to complete 60 items. Considering the mean and median item counts, it might be possible to develop measures of approximately 30-40 items that are as reliable and valid, and that require limited assessment time (i.e. 90-120 minutes). This supports claims for measures using data from multiple sources and potentially the use of pre-existing service records in order to streamline the assessment process (Essock et al., 2015, Schoenwald et al., 2011, 2013).

Domains of programme fidelity

Having discussed the common characteristics and approaches to fidelity assessment available to date, the final section of this review will tentatively address the question about the key domains of programme fidelity assessment. The classification of items into process, organizational structure, and delivery of care resemble Donabedian's

(1988) structure-process-outcome framework. As expected, most measures emphasized the relevance of structural features of the implemented programmes (i.e. operations, staff, team management, services, and interventions) whereas less tangible process features –such as the general model approach and specific client or practitioner actions– and outcomes were given less attention. This may be because of the complexity involved in operationalising some process variables (e.g. service goals or 'mission', availability, penetration, model focus, choice, service user involvement) and the appropriate outcomes and monitoring procedures for these interventions (e.g. days in hospital, time to relapse, self-report symptom trackers, functional scales). Nevertheless, some attempts have been made to include all three axes of Donabedian's (1988) framework, namely the CORE CRT and the FACTS, albeit with different degrees of cost-effectiveness.

Finally, item-level analysis could also shed some light on areas that may have been systematically overlooked in fidelity assessment. Using an implementation research framework (Proctor et al., 2009), most measures included features of policy and administration (e.g. philosophy, mission, availability, system involvement, and service-user choice), team operations (e.g. team approach, staff roles, caseload size, client to staff ratio, and multiagency integration), and intervention strategies (e.g. treatment tailoring, assertive outreach, mental health interventions, comprehensiveness of services, and crisis services). However, as stated above, little attention was paid to important implementation strategies such as training and consultation, and programme evaluation; the only exceptions were the CIMHRRS, FACTS, and IDDT measures. What is more, supervision –an important tool to support clinician adherence– is only mentioned by proxy (e.g. “team leader provides supervision”) on the IPS-25 and DDT. Similarly, none of the measures included items on continued professional development (CPD). In

relation to outcomes, more attention was given to client outcomes than to implementation and service outcomes. This was a little surprising, given that both training and outcome monitoring are important to ensuring and maintaining long-term therapist adherence and –arguably– programme fidelity (Carroll et al. 2007; Proctor et al., 2009). It is possible that although these measures might be picking up reliable cross-sectional indicators of fidelity (e.g. clinical practices, services provided, treatment model) they might not be identifying potential sources of model fidelity drift (e.g. training, CPD, supervision, staff turnover, outcome measures, process reviews, service-user feedback) (Forgatch & DeGarmo, 2011; Gearing et al., 2011).

Clinical implications

This systematic review is an updated and focused attempt at understanding how programme fidelity applies specifically to multi-component interventions in the context of severe mental health difficulties. Its exhaustive search strategy allowed the identification of a wide range of measures that might enhance the review from Bond and colleagues (2000a) fidelity toolkit. Findings suggest that, although programme fidelity has become a greater part of the general implementation agenda, there are still insufficient measures available and limited clarity as to (a) what they should measure and (b) how to do so effectively and efficiently. What is more, the relative lack of psychometric data makes it difficult to establish the quality of the measures. On a service level, this might translate to incomplete operational policies, inaccurate performance indexes, and ineffective training and development strategies. On a clinical level, however, the lack of reliable and valid fidelity checks could potentially lead to a lack of understanding of the active ingredients of a model and, consequently, poorer outcomes. Finally, results from the domain analysis might provide insights for the

typology of programme fidelity measures and help identify new areas to be included, such as items related to continued professional development, and a greater emphasis on outcomes.

Strengths and limitations

Although this systematic review expands our understanding of the content, characteristics and psychometric properties of programme fidelity measures, it has some limitations. The first limitation is that the definition of severe mental health difficulties was potentially too broad (i.e. including disorders other than schizophrenia, schizoaffective disorders and bipolar disorder, which are the most common diagnoses falling into this category; NIMH, 2019) and the definition of multi-component interventions too narrow (i.e. only including those with an active mental health component). As previously discussed, the former led to a quite complex search strategy (Appendix A), and the latter to the exclusion potentially helpful interventions. This could have been solved by a simpler definition of severe mental health difficulties or a focus on more specific diagnostic groups; however, when doing so our search strategy could not identify many of our seminal references. The final definitions used, nevertheless, were quite rigorous given that it was based on recent literature on the subject and was reviewed by an expert panel.

The second limitation was the single-rater screening procedure. Unfortunately, the search results were unmanageable to double-screen given the time constraints of the project. Likewise, there was no rigorous risk of bias assessment given the lack of published standards for this specific type of searches (i.e. service-level measures). These limitations were partly addressed by following relevant COSMIN recommendations (Mokkink et al., 2018), ensuring a high reliability between raters (both of which

collaborated in the scoping of the review), meeting after each stage to discuss any discrepancies and duplicates in the process, and including SP as a referee.

Finally, the item-level domain analysis was somewhat arbitrary and guided by the item themes themselves. This analysis was an attempt at organizing information from multiple measures to identify gaps in knowledge. This could have been done in a more systematic way, however, given the time frame and the complexity of the task this was not possible. Nevertheless, item sorting was done following Braun & Clarke's (2006) approach and domains were discussed with SP to improve their validity.

Conclusion

Programme fidelity is an important –yet somewhat neglected– aspect of implementation science. This neglect is partly due to a misunderstanding of the terminology, a lack of clarity of its key domains, and to the inherent complexity of measuring multiple interacting components in an effective and efficient way. Results from the present review suggests that existing programme fidelity measures of multi-component mental health interventions for severe mental health difficulties address primarily adult interventions. Arguably, there is some consistency in terms of measure design, assessment procedures, and scoring; however, there is greater variability in terms of the domains covered, with most emphasis given to structural features of service provision. Finally, a scarcity of validation studies in the field was noted, resulting in an inability to establish reliable benchmarks. Failing to addressing these shortcomings may result in services and clinicians drifting away from their models and, consequently, to unsustainable implementation efforts.

References

Aarons, G., Hurlburt, M., & Horwitz, S. (2011). Advancing a conceptual model of

evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health Services Research*, 38(1), 4-23. DOI:

10.1007/s10488-010-0327-7

Becker, D., & Drake, R. (1993). *A Working Life-the Individual Placement and Support (IPS) Program*. Concord, NH: New Hampshire-Dartmouth Psychiatric Research Center.

Becker, D., Smith, J., Tanzman, B., Drake, R., & Tremblay, T. (2001). Fidelity of supported employment programs and employment outcomes. *Psychiatric Services*, 52(6), 834-836.

Bloch, R., Saed, S., Rivard, J., & Rausch C. (2006). Lessons learned in implementing evidence-based practices: Implications for psychiatric administrators. *Psychiatric Quarterly*, 77, 309–318. DOI: 10.1007/s11126-006-9016-9

Bond, G., Becker, D., & Drake, R. (2011). Measurement of fidelity of implementation of evidence-based practices: Case example of the IPS Fidelity Scale. *Clinical Psychology: Science and Practice*, 18(2), 126-141. DOI: 10.1111/j.1468-2850.2011.01244.x

Bond, G., Becker, D., Drake, R., & Vogler, K. (1997). A fidelity scale for the individual placement and support model of supported employment. *Rehabilitation Counseling Bulletin*, 40, 265-284.

Bond, G., Campbell, K., Evans, L., Gervery, R., Pascaris, A., Tice, S., Del Bene, D., & Revell, G. (2002). A scale to measure quality of supported employment for persons with severe mental illness. *Journal of Vocational Rehabilitation*, 17(4), 239-250.

- Bond, G., & Drake, R. (2017). New directions for psychiatric rehabilitation in the USA. *Epidemiology and Psychiatric Sciences*, 26, 223–227. DOI: 10.1017/S2045796016000834
- Bond, G., Drake, R., & Becker, D. (2008). An update on randomized controlled trials of evidence-based supported employment. *Psychiatric Rehabilitation Journal*, 31, 280–290.
- Bond, G., Drake, R., Becker, D., & Noel, V. (2016). The IPS learning community: a longitudinal study of sustainment, quality, and outcome. *Psychiatric Services*, 67(8), 864–869. DOI: 10.1176/appi.ps.201500301
- Bond, G., Williams, J., Evans, L., Salyers, M., Kim, H., Sharpe, H., & Leff, H. (2000). *Psychiatric rehabilitation fidelity toolkit*. Cambridge, MA: Human Services Research Institute.
- Bond, G., Peterson, A., Becker, D., & Drake, R. (2012). Validation of the Revised Individual Placement and Support Fidelity Scale (IPS-25). *Psychiatric Services*, 63, 758–763.
- Bond, G., Picone, J., Mauer, B., Fishbein, S., & Stout, R. (2000b). The quality of supported employment implementation scale. *Journal of Vocational Rehabilitation*, 14(3), 201–212.
- Bond, G., & Salyers, M. (2004). Prediction of outcome from the Dartmouth assertive community treatment fidelity scale. *CNS spectrums*, 9(12), 937–942.
- Bond, G., Vogler, K., Resnick, S., Evans, L., Drake, R., & Becker, D. (2001). Dimensions of supported employment: Factor structure of the IPS Fidelity Scale. *Journal of Mental Health*, 10(4), 383–393. DOI: 10.1080/09638230124258

- Borelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, 71(S1), S52-S63. DOI: 10.1111/j.1752-7325.2011.00233.x
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Bruns, E., Burchard, J., Suter, J., Leverentz-Brady, K., & Force, M. (2004). Assessing fidelity to a community-based treatment for youth: The Wraparound Fidelity Index. *Journal of Emotional and Behavioral Disorders*, 12(2), 79–89. DOI: 10.1177/10634266040120020201
- Bruns, E., Leverentz-Brady, K., & Suter, J. (2008). Is it wraparound yet? Setting fidelity standards for the wraparound process. *Journal of Behavioral Health Services and Research*, 35, 240–252. DOI: 10.1007/s11414-008-9109-3
- Bruns, E., Suter, J., Force, M., & Burchard, J. (2005). Adherence to wraparound principles and association with outcomes. *Journal of Child and Family Studies*, 14(4), 521-534. DOI: 10.1007/s10826-005-7186-y
- Burchard, J., Bruns, E., & Burchard, S. (2002). The wraparound approach. In B. Burns & K. Hoagwood (Eds.), *Community treatment for youth: Evidence-based interventions for severe emotional and behavioral disorders*. Oxford: Oxford University Press.
- Cabaniss, D., Wainberg, M., & Oquendo, M. (2015). Commentary: Evidence-based psychosocial interventions: Novel challenges for training and implementation. *Depression and Anxiety*, 32:802-804. DOI: 10.1002/da.22437
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1),

40. DOI: 10.1186/1748-5908-2-40

Corbett, K., Thompson, B., White, N., & Taylor, M. (1991). Process evaluation in the community intervention trial for smoking cessation (COMMIT). *International Quarterly of Community Health Education*, 11, 291-309.

Cross, W., & West, J. (2011). Examining implementer fidelity: Conceptualizing and measuring adherence and competence. *Journal of Child Services*, 6(1): 18–33. DOI: 10.5042/jcs.2011.0123.

Donabedian, A. (1988). The quality of care: how can it be assessed?. *Journal of the American Medical Association*, 260(12), 1743-1748.

Drake, R., & Deegan, P. (2008). Are Assertive Community Treatment and Recovery Compatible? Commentary on “ACT and Recovery: Integrating Evidence-based Practice and Recovery Orientation on Assertive Community Treatment Teams”. *Community Mental Health Journal*, 44, 75–77. DOI 10.1007/s10597-007-9120-9

Drake, R., & Essock, S. (2009). The science-to-service gap in real-world schizophrenia treatment: The 95% problem. *Schizophrenia Bulletin*, 35(4), 677-678. DOI: 10.1093/schbul/sbp047

Drake, R., Goldman, H., Leff, H., Lehman, A., Dixon, L., Mueser, K., & Torrey, W. (2001). Implementing evidence-based practices in routine mental health service settings. *Psychiatric Services*, 52(2), 179-182. DOI: 10.1176/appi.ps.52.2.179

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237-256.

Effland, V., Walton, B., & McIntyre, J. (2011). Connecting the dots: Stages of implementation, wraparound fidelity and youth outcomes. *Journal of Child and*

Family Studies, 20(6), 736-746. DOI: 10.1007/s10826-011-9541-5

- Essock, S., Nossel, I., McNamara, K., Bennet, M., Buchanan, R., Kreyenbuhl, J., Mendon, S., Goldman, H., & Dixon, L. (2015). Practical monitoring of treatment fidelity: Examples from a team-based intervention for people with early psychosis. *Psychiatric Services*, 66(7), DOI: 10.1176/appi.ps.201400531
- Faw, L. (2003). *Multidimensional fidelity evaluation in a residential program for adolescent substance abuse* (Doctoral Dissertation). Retrieved from EBSCOhost, ProQuest Information & Learning. 64: 3557-3557.
- Forgatch, M., & DeGarmo, D. (2011). Sustaining fidelity following the nationwide PMTO implementation in Norway. *Prevention Science*, 12(3), 235-246. DOI: 10.1007/s11121-011-0225-6
- Fowler, D., Garety, P., Kuipers, E. (1995). *Cognitive-Behaviour Therapy for Psychosis: Theory and Practice*. Chichester, UK: Wiley.
- Gearing, R., El-Bassel, N., Ghesquierre, A., Baldwyn, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, 31, 79-88. DOI: 10.1016/j.cpr.2010.09.007
- Gresham, F., Gansle, K., Noell, G., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review*, 22, 254-272
- Harrison, J., Curtis, A., Cousins, L., & Spybrook, J. (2017). Integrated dual disorder treatment implementation in a large state sample. *Community Mental Health Journal*, 53(3), 358-366. DOI: 10.1007/s10597-016-0019-1

Hazelden Foundation (2016). *Severe Mental Illness Defined by Duration and Disability*.

Retrieved from http://www.bhevolution.org/public/severe_mental_illness.page

Henggeler, S., Rodick, J., Borduin, C., Hanson, C., Watson, S., & Urey, J. (1986).

Multisystemic treatment of juvenile offenders: Effects on adolescent behavior and family interactions. *Developmental Psychology*, 22, 132-141.

Higgins, J., & Deeks, J. (2008). Chapter 7: Selecting studies and collecting data. In J. T.

Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions*. Chichester: John Wiley & Sons.

Higgins, J., & Green, S. (2011). *Cochrane handbook for systematic reviews of*

interventions version 5.1.6 [updated September 2018]. The Cochrane Collaboration. Retrieved from <http://handbook.cochrane.org>

Johnsen, M., Samberg, L., Calsyn, R., Blasinsky, M., Landow, W., & Goldman, H.

(1999). Case management models for persons who are homeless and mentally ill: The ACCESS demonstration project. *Community Mental Health Journal*, 35(4), 325-346. DOI: 10.1023/A:1018761807225

Johnson, R. (2011). *Examining the interrater reliability of the comprehensive inventory*

of mental health and recovery and rehabilitation services (CIMHRRS). Theses, Dissertations, and Student Research: Department of Psychology. University of Nebraska. Retrieved from: <http://digitalcommons.unl.edu/psychdiss/20>

Kernan, J. (2014). Measuring wraparound fidelity to make quality improvements.

Community Mental Health Journal, 50(8), 903-908. DOI: 10.1007/s10597-014-9714-y

Kidd, S., George, L., O'Connell, M., Sylvestre, J., Kirkpatrick, H., Browne, G., &

- Thabane, L. (2010). Fidelity and recovery-orientation in assertive community treatment. *Community Mental Health Journal, 46*(4), 342-350. DOI: 10.1007/s10597-009-9275-7
- Latimer, E. (1999). Economic impacts of assertive community treatment: a review of the literature. *The Canadian Journal of Psychiatry, 44*(5), 443-454.
- Lloyd-Evans, B., Bond, G., Ruud, T., Ivanecka, A., Gray, R., Osborn, D., Nolan, F., Henderson, C., Mason, O., Goater, N., Kelly, K., Ambler, G., Morant, N., ONyett, S., Lamb, D., Fahmy, S., Brown, E., Paterson, B., Sweeney A., Hindle, D., Fullarton, K., Frerichs, J., & Johnson, S. (2016). Development of a measure of model fidelity for mental health Crisis Resolution Teams. *BMC psychiatry, 16*(1), 427. DOI: 10.1186/s12888-016-1139-4
- McGowan, J., Sampson, M., Salzwedel, D., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of Clinical Epidemiology, 75*, 40-60. DOI: 10.1016/j.clinepi.2016.01.021
- McGrew, J., Bond, G., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Counseling and Clinical Psychology, 62*(4), 670–678.
- McGrew, J., & Griss, M. (2005). Concurrent and predictive validity of two scales to assess the fidelity of implementation of supported employment. *Psychiatric Rehabilitation Journal, 29*(1), 41. DOI: 10.2975/29.2005.41.47
- McGrew, J., White, L., Stull, L., & Wright-Berryman, J. (2013). A comparison of self-reported and phone-administered methods of ACT fidelity assessment: A pilot study in Indiana. *Psychiatric Services, 64*(3), 272-276. DOI:

10.1176/appi.ps.001252012

- McGrew, J. H., Wilson, R. G., & Bond, G. R. (2002). An exploratory study of what clients like least about assertive community treatment. *Psychiatric Services, 53*(6), 761-763.
- McHugo, G., Drake, R., Whitley, R., Bond, G., Campbell, K., Rapp, C., Goldman, H., Lutz, W., & Finnerty, M. (2007). Fidelity outcomes in the national implementing evidence-based practices project. *Psychiatric Services, 58*(10), 1279-1284
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth, 4*, 83-86
- Mokkink, L., De Vet, H., Prinsen, C., Patrick, D., Alonso, J., Bouter, L., & Terwee, C. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1171-1179.
- Moncher, F., & Prinz, R. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*(3), 247-266.
- Monroe-DeVita, M., Teague, G., & Moser, L. (2011). The TMACT: a new tool for measuring fidelity to assertive community treatment. *Journal of the American Psychiatric Nurses Association, 17*(1), 17-29. DOI: 10.1177/1078390310394658
- Mueser, K., Noordsy, D., Drake, R., & Fox, L. (2003). *Integrated treatment for dual disorders: A guide to effective practice*. New York: Guilford Press.
- National Institute of Mental Health (February, 2019). *Mental illness*. Retrieved from <https://www.nimh.nih.gov/health/statistics/mental-illness.shtml>
- Nugter, M., Engelsbel, F., Bähler, M., Keet, R., & van Veldhuizen, R. (2016). Outcomes of FLEXIBLE Assertive Community Treatment (FACT) implementation: A prospective real life study. *Community Mental Health Journal, 52*(8), 898-907.

DOI: 10.1007/s10597-015-9831-2

- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P., Cooper, J., Eaton, J., Herrman, H., Herzallah, M., Huang, Y., Jordans, M., Kleinman, A., Medina-Mora, M., Morgan, E., Niaz, U., Omigbodun, O., Prince, M., Rahman, A., Saraceno, B., Sarkar, B., De Silva, M., Singh, I., Stein, D., Sunkel, C., & Unützer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553-1598.
- Pentz, M., Trebow, E., Hansen, W., MacKinnon, D., Dwyer, J., Johnson, C., Flay, B., Daniels, S., & Cormack, C. (1990). Effects of program implementation on adolescent drug use behavior: The Midwestern Prevention Project (MPP). *Evaluation Review*, 14, 264-289.
- Perepletchikova, F., Treat, T., & Kazdin, A. (2007). Treatment integrity in psychotherapy research: analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75(6), 829.
- Petrakis, M., Penno, S., Selvendra, A., Laxton, S., Doidge, G., & Castle, D. (2011). Fidelity to clinical guidelines using a care pathway in the treatment of first episode psychosis. *Journal of Evaluation in Clinical Practice*, 17, 722-728.
DOI:10.1111/j.13652753.2010.0 1548.x
- Popay, J., Robert, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. *A product from the ESRC methods programme Version 1*, b92.
- Prinsen, C., Mokkink, L., Bouter, L., Alonso, J., Patrick, D., de Vet, H., & Terwee, C.

- (2016). COSMIN guideline for systematic reviews of outcome measurement instruments. *Quality of Life Research*, 25, 21-21.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Giffey, R., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65-76.
- Public Health England (September, 2018). *Research and analysis Severe mental illness (SMI) and physical health inequalities: briefing*. Retrieved from <https://www.gov.uk/government/publications/severe-mental-illness-smi-physical-health-inequalities/severe-mental-illness-and-physical-health-inequalities-briefing#fn:1>
- Pullmann, M., Bruns, E., & Sather, A., (2013). Evaluating fidelity to the wraparound service model for youth: application of item response theory to the Wraparound Fidelity Index. *Psychological Assessment*, 25(2), 583. DOI: 10.1037/a0031864
- Rollins, A., McGrew, J., Kukla, M., McGuire, A., Flanagan, M., Hunt, M., Leslie, D., Collins, L., Wright-Berryman, J., Hicks, L., & Salyers, M. (2016). Comparison of assertive community treatment fidelity assessment methods: Reliability and validity. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(2), 157-167. DOI: 10.1007/s10488-015-0641-1
- Rosenheck, R., Kaspro, W., Frisman, L., & Liu-Mares, W. (2003). Cost-effectiveness of supported housing for homeless persons with mental illness. *Archives of General Psychiatry*, 60(9), 940-951.
- Salyers, M., Bond, G., Teague, G., Cox, J., Smith, M., Hicks, M., & Koop, J. (2003). Is it ACT yet? Real-world examples of evaluating the degree of implementation for

- assertive community treatment. *The Journal of Behavioral Health Services & Research*, 30(3), 304-320. DOI: 10.1007/BF02287319
- Salyers, M., & Tsembris, S., (2007). ACT and Recovery: Integrating Evidence-Based Practice and Recovery Orientation on Assertive Community Treatment Teams. *Community Mental Health Journal*, 43(6), DOI: 10.1007/s10597-007-9088-5
- Sanetti, L., & Kratochwill, T. (2009). Treatment integrity assessment in the schools: An evaluation of the Treatment Integrity Planning Protocol. *School Psychology Quarterly*, 24(1), 24.
- Santacroce, S., Maccarelli, L., & Grey, M. (2004). Intervention fidelity. *Nursing Research*, 52(1), 63-66.
- Schoenwald, S., & Garland, A. (2013). A review of treatment adherence measurement methods. *Psychological Assessment*, 25(1), 146–156. DOI:10.1037/a0029715.
- Schoenwald, S., Garland, A., Chapman, J., Frazier, S., Sheidow, A., & Southam-Gerow, M. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(1), 32-43. DOI: 10.1007/s10488-010-0321-0
- Stein, L., & Test, M. (1980). Alternatives to mental hospital treatment Part I: Conceptual model treatment program and clinical evaluation. *Archives of General Psychiatry*, 37, 392–397. DOI: 10.1007/s10597-015-9831-2
- Teague, G., Bond, G., & Drake, R. (1998). Program fidelity in assertive community treatment: development and use of a measure. *American Journal of Orthopsychiatry*, 68(2), 216-233. DOI: 10.1037/h0080331
- Teague, G., Drake, R., & Ackerson, T. (1995). Evaluating use of continuous treatment teams for persons with mental illness and substance abuse. *Psychiatric Services*,

46(7): 689-695.

- Teague, G., Mueser, K., & Rapp, C. (2012). Advances in fidelity measurement for mental health services research. *Psychiatric Services*, 63(8), 765-771. DOI: 10.1176/appi.ps.201100430
- Torrey, W., Bond, G., McHugo, G., & Swain, K. (2012). Evidence-based practice implementation in community mental health settings: The relative importance of key domains of implementation activity. *Administration and Policy in Mental Health*, 39, 353-364. DOI 10.1007/s10488-011-0357-9
- Torrey, W., Finnerty, M., Evans, A., & Wyzik, P. (2003). Strategies for leading the implementation of evidence-based practices. *Psychiatric Clinics of North America*, 26(4), 883-897. DOI: 10.1016/S0193-953X(03)00067-4
- Vogler, K. (1998). *A fidelity study of the Indiana Supported Employment Model for individuals with severe mental illness*. Unpublished dissertation, Indiana University-Purdue University Indianapolis, Indianapolis, IN.
- Waghorn, G. (2009). Measuring the fidelity of supported employment for people with severe mental illness. *Australian Occupational Therapy Journal*, 56, 367–368. DOI: 10.1111/j.1440-1630.2009.00809.x
- Walker, J., & Bruns, E. (2013). *Wraparound fidelity assessment system*. *Wraparound evaluation and research team*. Retrieved from <http://depts.washington.edu/wrapeval/index.html>
- Wheeler, C., Lloyd-Evans, B., Churchard, A., Fitzgerald, C., Fullarton, K., Mosse, L., Paterson, B., Galli Zugaro, C., & Johnson, S. (2015). Implementation of the Crisis Resolution Team model in adult mental health settings: a systematic review. *BMC Psychiatry*, 15, 74. DOI: 10.1186/s12888-015-0441-x

Wilson, D. & Crisanti, A. (2009). Psychometric properties of the dual-disorder treatment fidelity scale: inter-rater reliability and concurrent validity. *Community Mental Health Journal*, 45(3): 171-178. DOI: 10.1007/s10597-008-9167-2

Winter, J. & Calsyn, R. (2000). The Dartmouth Assertive Community Treatment Scale (DACTS). A generalizability study. *Evaluation Review*, 24(3): 319-338. DOI: 10.1177/0193841X0002400304

Part 2: Empirical Paper

The Community Mental Health Fidelity Scale (CoMFideS):

A measure of programme fidelity of a social networks intervention for
severe mental health difficulties

Abstract

Open Dialogue (OD) is a multi-component therapeutic and organizational intervention for crisis and continuing community mental health care with a therapeutic focus on clients' social networks. The development and implementation of this model of care in the United Kingdom requires considerable contextual adaptations which need to be assessed to support effective implementation. Programme fidelity –the extent to which core components of an intervention are delivered as intended by an intervention protocol at all levels– is crucial for these adaptations. *Aims:* To develop and pilot a programme fidelity measure for community mental health services providing OD and 'care as usual' (CAU) or standard NHS crisis and community care. *Methods:* Measure structure, content, and scoring were developed and refined through an iterative process of discussion between the research team and OD experts. Measure was piloted in the 6 OD and 6 CAU services participating in a large-scale research programme. *Results:* Initial data suggests that the *Community Mental Health Fidelity Scale (CoMFideS)* is a potentially reliable and feasible measure of the fidelity of community mental health services and specific OD components of such services.

Introduction

Poor social networks have been associated with both the development and maintenance of mental health difficulties (Giacco et al., 2012). Interventions targeting social networks –such as the Open Dialogue approach (Seikkula et al., 1995)– might therefore help ameliorate mental health crises and reduce the likelihood of relapse. However, due to limited staff knowledge and skills, and a lack of continuity in the current NHS model of crisis and continued community care, such interventions are not currently provided (Razzaque & Wood, 2015; The Commission on Acute Adult Psychiatric Care, 2015). Further, the professional and contextual adaptations required to successfully and sustainably integrate Open Dialogue into the traditional NHS model of care require a consideration of the model’s “active components”.

As explored in the previous chapter, programme fidelity or the extent to which core components of an intervention are delivered as intended by a treatment protocol at all levels, is a useful approach to identifying the key components of an intervention (Borelli, 2011; Gearing et al., 2011; Santacroce, Maccarelli, & Grey, 2004). This chapter presents an empirical study aiming to develop, pilot, and implement a programme fidelity measure for the Open Dialogue approach in the NHS. The chapter begins with a brief description of Open Dialogue and the traditional NHS model of crisis and continuing community care in mental health. This is followed by an exploration of some of the challenges involved in integrating Open Dialogue to the provision of mental health services in United Kingdom, including the challenges in developing fidelity measures for said purposes. Next, the study aims and methods are outlined, paying special attention to the measure development process. Descriptive statistics and preliminary psychometric properties are then presented and discussed. Finally, the

strengths and limitations of the study are examined, as well as some clinical implications and future developments of the measure at hand.

The Open Dialogue approach

Open Dialogue (Seikkula et al., 1995) is a therapeutic approach and way of organising mental health services developed in Finland, which explicitly targets social networks. The aim of Open Dialogue is to promote a greater shared understanding of service users' problems, a greater sense of agency, collaborative decision making, and the network's mutual support in the long term (Seikkula, Alakare, & Aaltonen, 2001a, 2011; Seikkula et al., 1995, 2006). In theory, this is done through the enactment of the principles of (1) immediate help (i.e. within 24 hours), (2) social networks perspective, (3) flexibility and mobility (of modalities and service provision), (4) responsibility (in care coordination), (5) psychological continuity, (6) tolerance of uncertainty, and (7) dialogue and polyphony (Seikkula et al., 1995). In contrast to current models of care –in which families are rarely directly involved– Open Dialogue uses network meetings attended by family members, friends, and other professionals involved with the service user as the central means of intervention delivery (Lakeman, 2014; Seikkula, et al., 1995; Seikkula & Olson, 2003). Service users and their social network engage in shared decision making with healthcare professionals to agree on appropriate pharmaceutical, psychological, or social interventions (Olson, Seikkula, & Ziedonis, 2014).

Open Dialogue is rooted on Gregory Bateson's systemic theories of psychosis (Bateson, Jackson, Haley, & Weakland, 1956) and Mikhail Bakhtin's (1981) ideas on discourse analysis, specifically on the notions of polyphony (i.e. multiple voices) and dialogism (i.e. a live interaction where these voices engage from different –yet equally valid– perspectives) (Seikkula & Olson, 2003). As such, Open Dialogue is inherently

relational and focuses on the interpersonal interactions within the social network, as it is there where the speakers are embedded and ‘truth’ is negotiated (Mikes-Liu, 2015; Seikkula & Trimble, 2005).

The NHS model of crisis and continuing community care in mental health

The National Health Service is currently facing significant problems in providing care and support for people with severe mental health difficulties. Pathways through care are poorly developed and increasingly fragmented (NHS Confederation, 2016; The Kings Fund, 2016). This is in part a consequence of the functional model of mental health care, where care is often provided by several different teams, each with their own criteria for acceptance (Morton & Norman-Nott, 2019). Traditional NHS crisis and continuing community care services for people experiencing severe mental health difficulties (hereafter referred to as ‘care as usual’ or CAU) consist primarily of crisis resolution and home treatment teams (CRTs) and community mental health teams (CMHTs). As an alternative to hospitalization, these multidisciplinary teams – typically conformed by psychiatrists, mental health nurses, social workers, and support workers – provide intensive assessment, care, and support in patients’ homes (Jethwa, Galappathie, & Hewson, 2007; Weisman, 1989). CRTs and CMHTs often acknowledge and may attempt to work with the social network of a person in crisis; however, their brief and functional nature and the pressures on service resources make this form of ongoing network-oriented care a challenging endeavour (Razzaque & Wood, 2015).

Despite the promise shown in randomised controlled trials (RCTs) (Johnson et al., 2005a, 2005b, Lloyd-Evans et al., 2014, 2019), questions have been raised on whether CAU might be decreasing in effectiveness (Jacobs & Barrenho, 2011; Johnson et al., 2005a, 2005b). For instance, a systematic review on CRTs by Wheeler and

colleagues (2015) suggested this might be due to a considerable atrophy of its key functions, with many services offering limited home visits outside of office hours and only 50% of services providing post-hospital discharge care. It is important to ask whether this possible decrease in quality of community-based services can be explained by lack of resources or if organizational problems, such as staff competences, roles, care pathways, or fidelity to a model, may also be contributory factors.

Open Dialogue in United Kingdom

The development of an integrated Open Dialogue approach to the provision of mental health services offers the possibility of an effective alternative to the current functional model of care in the United Kingdom. Preliminary evidence from Finland suggests that Open Dialogue could be more effective than CAU in reducing relapse and the use of antipsychotic medication (Bergström et al., 2018; Hartman & De Coursey, 2015; Rasinkangas & Lehtinen, 2003; Razzaque & Stockman, 2015; Seikkula, Alakare, & Aaltonen, 2001b). Additionally, the development of such an approach would equip mental health staff with the skills necessary to engage service users and their families across the broad spectrum of care needs (Holmesland, Seikkula, & Hopfenbeck, 2014). Although promising, nonetheless, there is no high-quality evidence to support a NHS-wide adoption of this model of care.

In 2016, Pilling and colleagues secured a 5-year NIHR programme grant (RP-PG-0615-20021) to assess whether Open Dialogue –when integrated within standard NHS mental health services for adults in crisis– improves the clinical and cost-effectiveness of traditional crisis and continuing community mental health care (i.e. CRTs and CMHTs). The challenge nonetheless is that Open Dialogue is both a psychological as well as an organizational intervention that requires the reorganization

of entire care pathways. Transferring such a model of care from one culture to another requires considerable contextual adaptations that could undermine structural (i.e. organizational) and process (i.e. therapeutic) components of the original model (Gonzalez Castro, Barrera, & Martinez, 2004). In fact, several international Open Dialogue implementation programmes (e.g. Gillard et al., 2015; Johnson, 2013; Pocobello & Salamina, 2015) have noted that the organizational change is such, that staying faithful to the seven principles of Open Dialogue (Seikkula et al., 2006) and the 12 key elements of dialogic practice (Olson et al., 2014) has encountered significant obstacles.

Programme fidelity measurement

The previous chapter explored the complexity behind understanding how the different components of an intervention influence outcomes in multi-component interventions such as Open Dialogue or community mental health interventions. This is arguably due to their large number of interacting contextual, organizational, and service-level components (Teague, Mueser, & Rapp, 2012; Wheeler et al., 2015). Therefore, developing measures that can effectively and efficiently identify said components is becoming a major endeavour for the field of implementation research (Proctor et al., 2009; Schoenwald et al., 2011; Teague et al., 2012).

Literature suggest that programme fidelity measures should involve (1) an evidence-based, comprehensive, and multimodal approach to assessment, (2) with clearly and objectively operationalised components stemming from a coherent and comprehensive theory of change, and (3) easily-available data from the relevant stakeholders (Essock et al., 2015; Schoenwald, & Garland, 2013). Existing measures for multi-component interventions for people with severe mental health difficulties are

somewhat consistent in terms of measure design, assessment procedures, and scoring; yet, as evidenced by findings from our systematic review, there is no consensus about which domains need to be included. Donabedian (1988) suggested a structure-process-outcome framework to fidelity evaluation; however, as illustrated in our systematic review, most measures emphasise on structural features of service provision (e.g. operations, staffing, or services provided) but tend to neglect important process and outcome features relevant to the therapeutic model.

A few efforts have been made to establish appropriate fidelity measures for CAU. The CORE CRT (Lloyd-Evans et al., 2016) is the most robust and validated measure to date for crisis services (please refer to Chapter 1, p. 29 for an in-depth description). Similarly, some efforts have been made for assessing fidelity to Open Dialogue, such as the ‘10 Organizational Criteria of Open Dialogue’ (Ziedonis et al., 2015) which highlights some important process features of the approach (e.g. routine network meetings, connecting services, dialogic practice) but has not yet been validated. Since Open Dialogue is not only a therapeutic model but also –potentially– a more successful way of organizing care, it is important to identify not only the clinically-relevant (i.e. process) features but also the structural and organizational ones that characterise the approach. In other words, if Open Dialogue is to be successfully implemented and integrated to the traditional NHS model of crisis and continued community care, it is quintessential to develop a programme fidelity measure to inform the implementation framework for its delivery in a way that is faithful not only to the original Finnish model, but also fit for its incorporation into the NHS.

Study aims

The key goal of the present study was to develop, pilot, and implement a reliable

programme fidelity measure for its use on the ODDESSI programme, to characterise the quality of both NHS crisis and continuing community care (CAU) and high-quality Open Dialogue (OD) practice. If successful, this measure would help establish whether:

1. NHS services, once reorganised into the Open Dialogue model of care, can deliver OD with sufficient fidelity to its core principles while also operating effectively within CAU principles.
2. It is possible to distinguish OD services from CAU based on their model of work.
3. There are any differences in implementation between each model's teams.

Method

Setting

This study was part of the NIHR ODDESSI (Open Dialogue: Development and Evaluation of a Social Network Intervention for Severe Mental Illness) programme grant. The ODDESSI trial consists of five work-packages oriented towards defining, implementing, and evaluating Open Dialogue services in the NHS (for more information about the ODDESSI programme see NIHR grant RP-PG-0615-20021).

Participants

ODDESSI is a cluster-RCT consisting of 28 trial clusters from five NHS Trusts. Trial clusters were sequentially allocated to OD or CAU in a 1:1 ratio. Open Dialogue services worked flexibly but were not 24/7 services; as such, external services (e.g. inpatient care or designated crisis teams) provided additional support and were invited to network meetings to ensure continuity of care. CAU teams were routine crisis care and continuing community care services (i.e. CRTs and CMHTs), which included the provision of care coordination and a range of psychosocial, pharmacological, and

psychological interventions. Randomisation was stratified by catchment area and balanced by two continuous cluster-level covariates, namely GP list (i.e. caseload) size and deprivation rate. The 6 OD teams participating in the ODDESSI trial were compared against their local 6 CAU counterparts.

Ethical Approval

This study received ethics approval from the Health Research Authority under reference number 18/LO/0026 (Appendix H). Participants for this study were only staff members. No service-users participated in this study and no personal or confidential information was solicited.

Trial design

Although this specific study was relevant to all work packages of the ODDESSI trial, it was embedded in the first work package as part of the feasibility stage (WP1). WP1 addresses the development of the intervention, its feasibility, and its acceptability to staff, service users, carers, and the wider organisational context in which it is implemented. Additionally, the NIHR shared their concern that –in order to draw meaningful conclusions from the outcomes– the trial needed to be able to compare OD teams against high quality CAU. Developing a programme fidelity measure is one way of addressing these concerns. Since dialogic practice and CAU both have systemic principles at their core (Seikkula & Olson, 2003), it was important to develop a measure that could encompass both models, while simultaneously recognising features of high-quality CAU.

Measure development

The *Community Mental Health Fidelity Scale* (CoMFideS) was developed following a stepwise approach (Bond et al., 2000; Holmbeck & Devine, 2009), based on

existing measures, findings from our systematic review (see Chapter 1), and a series of discussions with experts (Figure 1).

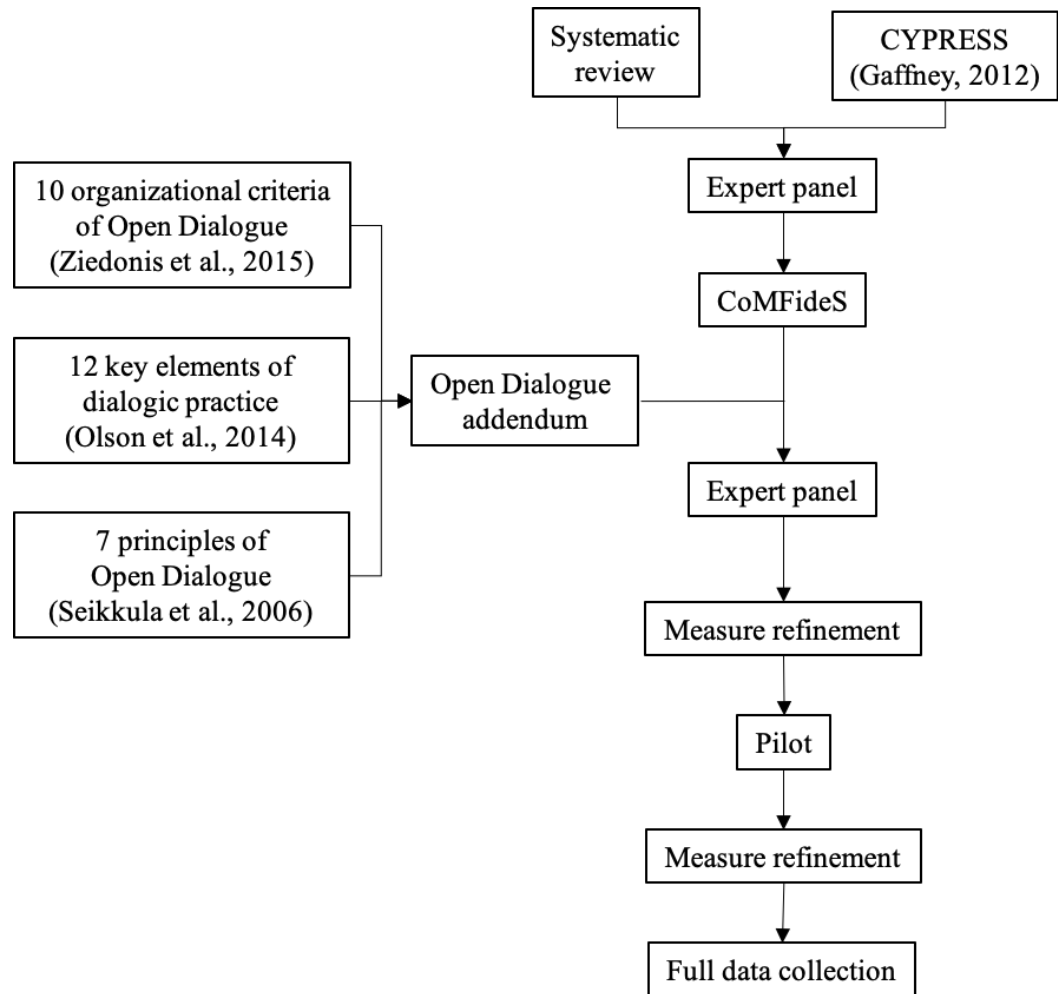


Figure 1. Community Mental health Fidelity Scale (CoMFideS) development process.

Defining the content and scope of the measure. The initial content, method of delivery, and scoring process of the CoMFideS were based on the ‘Children and Young People – Resource, Evaluation and Systems Schedule’ (CYPRESS) (Gaffney, 2012; Pilling, Butler, Gaffney, & Fonagy, 2012) and findings from our systematic review. CYPRESS was developed for the Systemic Therapy for At Risk Teens (START) RCT to

characterise services delivering multisystemic therapy and management as usual for young people with complex presentations. CYPRESS was designed based on existing measures and literature in the field; it captures key elements of effective implementation efforts (e.g. coherent theoretical basis, high programme fidelity, qualified staff, sustained approach, etc.) across three levels of service delivery: service characteristics, team operations, and delivery of interventions. The promising results from Cressida Gaffney's (2012) doctoral dissertation study suggested that CYPRESS could be a robust measure for service characterisation.

Drawing on the CYPRESS, findings from our systematic review (particularly the identified fidelity domains, p. 36), and Donabedian's (1988) structure-process-outcome framework, Mauricio Alvarez (MA; Trainee Clinical Psychologist) and Professor Stephen Pilling (SP; Clinical Psychologist with expertise in evidence-based practice and experience of measure development) agreed to four broad key domains to assess: (1) service structure and culture, (2) access to and engagement with services, (3) delivery of care, and (4) community linkage and support. An initial list of items was drafted for each domain and then refined based on three factors: (1) a focus on adult mental health, (2) ability to encompass both OD and CAU, and (3) ability to identify high quality CAU.

Designing the measure. The refinement and detail of the outline measure was established through a series of meetings and discussions between MA and Dr Russell Razzaque (RR; Consultant Psychiatrist, grand co-applicant, and national training coordinator for the Open Dialogue diploma in the UK), Mark Hopfenbeck (MH; lead OD trainer), and Melissa Lotmore (ML, Trainee Clinical Psychologist responsible of refining a treatment adherence measure for OD practitioners). An iterative refinement process took place between October 2017 and January 2018, aimed at achieving an

acceptable level of face and content validity of measure items.

Open Dialogue fidelity. Another important goal of WP1 was to refine the OD protocol for its implementation across NHS sites. Consequently, the resulting measure needed to be able to recognise features specific to Open Dialogue in OD teams. Based on their expertise on Open Dialogue, RR and MH agreed that the items in the CoMFideS were not specific enough to the model and would thus need a supplementary section to ascertain these features. The outcome of the process was reviewed by MA and SP.

A similar item development process took place based on existing Open Dialogue literature (e.g. Olson et al., 2014; Seikkula et al., 2006), the ‘10 Organizational Criteria of Open Dialogue’ (Ziedonis et al., 2015), and the ODDESSI treatment protocol. The ODDESSI protocol was developed by the research team in collaboration with experts in the fields of Open Dialogue and CAU, alongside senior NHS staff and clinicians. The protocol set out the core functions of an OD team and included for each site: (a) key functions and organisational structure of OD teams; (b) the referral pathways, caseload capacity, and team composition of the OD services; and (c) the support, supervision, and governance arrangements for each site.

Given the complexity of Open Dialogue terminology, a series of discussions around the main theoretical principles (e.g. dialogism, transparency, polyphony, openness, self-work, and context) were arranged with Open Dialogue experts. The aim was to determine the best possible way to translate these abstract terms into objective and reliable service-level items that could be ascertained by raters not trained in Open Dialogue. Prof Jaako Seikkula (developer of Open Dialogue; University of Jyväskylä, Finland), Prof Douglas Ziedonis (developer of the “10 Organizational Criteria for Open Dialogue”; University of California San Diego, USA), and Bjarne Vind (PhD candidate

focusing on “Openness and context in Open Dialogue”, Skovvænget, Denmark) participated in the discussions. Five main programme fidelity components were operationalised: (1) transparency (i.e. all discussions about the service-user’s treatment plan take place while the service-user is present in the room); (2) self-disclosure (i.e. clinical staff members sharing, where relevant, their own lived experiences); (3) intervision (i.e. supervision) arrangements; (4) team self-work practices (i.e. practices that foster self-knowledge and self-development, such as ‘family of origin’; Hopfenbeck, 2015) and (5) Open Dialogue-specific training and continued professional development (CPD). An initial list of items was drafted for an OD addendum and then refined based on discussions between MA, RR, MH, ML and SP.

The Community Mental health Fidelity Scale (CoMFideS)

The above lead to the development of a 25-item *Community Mental health Fidelity Scale (CoMFideS)*. CoMFideS is a measure designed to describe the structure, functioning, pathways, community links, and delivery of care provided by all high-quality CAU team including those who may provide OD (Appendix I). Also based on the aforementioned discussions, the CoMFideS also includes a 7-item Open Dialogue Addendum focused on measuring the level of fidelity to Open Dialogue principles of care. CoMFideS is hence a measure of programme fidelity of both (a) standard NHS crisis and continued community care and (b) best practice in OD delivery (Table 1).

The first section of the CoMFideS is a front sheet where raters take note of different structural aspects of the sites under assessment, namely (1) team setup (i.e. standalone or integrated with other services), (2) number of employed staff (e.g. full time, temporary staff, turnover, roles available in the team), (3) caseloads (i.e. current, service maximum, individual maximum, individual average), (4) supervision and team

Table 1. Community Mental Health Fidelity Scale (CoMFideS) items

CoMFideS scale	
Service structure and culture (SSC)	
SSC1	Service ethos and comprehensiveness
SSC2	Staff training
SSC3	Supervision
SSC4	Staff roles
SSC5	Service capacity
SSC6	Routine outcome measurement
SSC7	Safety
SSC8	Service-user involvement in co-production
Access and engagement (AE)	
AE1	Access to the service
AE2	Providing information
AE3	Identification of support systems
AE4	Prompt action
AE5	Flexibility of response
AE6	Assertive engagement
Delivery of care (DC)	
DC1	Continuity of care
DC2	Establishing clinical meetings
DC3	Collaborative decision-making
DC4	Information sharing and communication
DC5	Service-user involvement in delivery of care
DC6	Coordination of care
Community linkage and support (CLS)	
CLS1	Service linkage
CLS2	Community links (Practitioner level)
CLS3	Community links (Support system)
CLS4	Caregivers' involvement and support
CLS5	Discharge and aftercare
Open Dialogue Addendum (OD)	
OD1	Transparency
OD2	Self-disclosure
OD3	Intervision frequency
OD4	Intervision content and structure
OD5	Team self-work
OD6	OD Training
OD7	OD Continued professional development

meeting arrangements, and (5) a diagrammatic or verbal description of the organizational structure of the team (e.g. access and referral pathways, associated services, and relevant teams or agencies).

The CoMFideS is then divided into four fidelity sections that assess the level of fidelity of mental health teams –regardless of their model of care– to high-quality crisis and continued community care (CAU):

1. *Service structure and culture (8 items)* is concerned with the service ethos and comprehensiveness of the model of care (i.e. a well-articulated and shared view of the model of care provided), staff training, supervision, staff roles, service capacity, routine outcome monitoring, safety practices (for service users and staff), and the degree to which service users get involved in service co-production (i.e. development and evaluation of services).

2. *Access and engagement (6 items)* is concerned with referral and treatment pathways, provision of information about the service (to referring agencies and service users), prompt action, the proactive and effective inclusion of the service-user’s support systems, flexibility of response (i.e. range of interventions available), and assertive engagement strategies.

3. *Delivery of care (6 items)* is concerned with continuity of care (i.e. the extent to which the same care coordinator is maintained throughout a care pathway), the way in which clinical appointments or network meetings are arranged and convened, collaborative decision making, information-sharing and communication practices, coordination of care practices, and the degree to which service users get involved in the provision of care.

4. *Community linkage and support (5 items)* is concerned with the extent to

which the service is embedded in the local community (i.e. the amount and strength of links to other services and support agencies) and practitioners' skilfulness in effectively using these resources (e.g. liaison, follow-up, inclusion to meetings) and engaging the service-user's own support systems (instead of deploying the services' own resources). This section also focuses on the extent to which caregivers are taken into consideration and supported throughout treatment and the nature and sustainability of discharge and aftercare practices.

Additionally, an *Open Dialogue addendum (7 items)* evaluates the extent to which OD teams support and enact specific service-level dialogic principles into practice. This section assesses how the principle of transparency is enacted in the team, the attitude of the team towards self-disclosure, the team's frequency and content of intervision, team self-work practices, the promotion of certified Open Dialogue training courses, and the presence of a coherent Open Dialogue CPD programme. This section is only relevant to OD teams.

Initial pilot and measure refinement. The CoMFideS was piloted in one OD and one CAU service from the lead Trust of the programme. The goal was to identify areas of improvement in the CoMFideS and assess the measure's acceptability; it also helped determine whether it captured the theorised distinctive features of Open Dialogue. For each team, two pairs of managers and practitioners (e.g. psychotherapists, counsellors, nurses, psychiatrists, peer support workers, etc.) were interviewed using the draft measure, followed by a brief discussion on its structure, content, and acceptability. MA and SP carried out the pilot interviews. SP lead the pilot interviews and MA took notes on the discussions, the content of the measure, and the process of administration. Initial percentages of inter-rater agreement were 83% for the interview to OD managers,

79% for OD clinicians and 94% for CAU managers and 84% for CAU clinicians. After discussion between raters, all ratings were agreed upon for all three interviews. Using the outcome from the pilot, the measure was once again refined and discussed with the expert panel (See Figure 1).

The CoMFideS Manual. Based on insights gathered throughout the measure development process and drawing from CYPRESS (Pilling et al., 2012), MA developed a manual for the CoMFideS (Appendix J). The manual includes a detailed description and rationale for each item alongside their respective behavioural anchors for scoring. The manual also includes a service documentation checklist and the interview schedules for both interviews (i.e. managers and practitioners).

Full data collection

Recruitment and setting up the interviews. With the final measure and manual ready for rollout, MA contacted the participating services to arrange visits for the interviews. As per the manual, initial contact with services was done via email correspondence with each Trust's research assistant. Research assistants received a description of the fidelity assessment study, its purpose, and a brief description of the measure and interview process. Two managers and two practitioners from each site were contacted by MA and were selected based on availability for interviews. All participants gave verbal consent for participating in the study. As per measure manual, service documentation (e.g. staffing, supervision, safeguarding, and operational policies) was also requested from each Trust. This was intended to gather service-level data prior to the interviews and identify areas where further information was necessary.

Conducting the interviews. Interviews were programmed to last no more than 60 minutes. The average time spent per interview was 46 minutes (range=35-57). For

each interview, MA and one of the five external raters (ML, EW, KC, SP, EM) visited each site to interview staff (see Appendix J for a description of the interview process). All five secondary raters received a copy of the measure manual and received a training on the use of CoMFideS, which included discussions of each item and rating examples.

Given the pilot nature of this project it was not possible to fully nest or cross raters, which led to an ill-structured measurement design (Putka, Le, McCloy, & Diaz, 2008). MA led all interview sessions and second raters were determined based on availability on the agreed dates. Before the beginning of each interview, participants gave verbal consent for being recorded using an encrypted and password protected recorder. The purpose of the recording was to allow for full reliability testing later in the trial. No personal data was requested or elicited and all recordings were transferred to the programme's encrypted drive at the earliest chance.

All 25 CoMFideS items and 7 items on the OD-addendum were included in both CAU and OD interviews. The reason for using the OD-addendum on CAU teams was to establish whether, in fact, these items were relevant *primarily* to the Open Dialogue approach.

Agreeing final ratings. Once each interview session was completed, both raters reviewed their individual scores separately to confirm no information had been missed. Each item was then jointly reviewed to identify and record disagreements and reach a consensus on the final score. If any disagreement could not be resolved, SP acted as a referee.

Scoring. All items of the CoMFideS are rated on a 4-point behaviourally-anchored Likert scale. As discussed in the previous chapter (p. 43), Bond and colleagues (1997) identified that this method could potentially achieve the highest level of precision

possible for programme fidelity ratings, rather than the traditional 5-point Likert approach. A 4-point scale is thought to help avoid ‘neutral’ responses (i.e. scores of ‘3’) to more abstract items (e.g. ‘assertive engagement’, ‘collaborative decision making’, ‘transparency’, ‘self-disclosure’, and ‘team self-work’). As per manual, an item obtained a score of 1 when the features at hand were not present or there was insufficient evidence of their enactment in the team’s way of functioning. Contrarily, items with a score of 4 indicated that the features were enacted or fully carried out and with no shortcomings or inconsistencies across the team. Each section obtained an average score of its composite items.

Cut-off scores. According to the systematic review most programme fidelity measures use a three-tier cut-off structure (see Appendix D); however –after comparing the Quality of Supported Employment Implementation Scale (QSEIS) and the 15-item Individual Placement and Support (IPS) Fidelity Scale– Bond and colleagues (2002) advised that a 4-tier structure might allow for a finer discrimination between programmes. With this in mind and providing (1) this is the first fidelity measure developed for Open Dialogue in the NHS, and (2) that there are no pre-existing criteria for what constitutes a ‘good’ standard of CAU care, nor of Open Dialogue fidelity, we considered 4 fidelity gradations: an average score equal or above 3.40 was considered ‘very good’; scores between 2.80 and 3.39 as ‘good’; scores between 2.40 and 2.79 as ‘acceptable’; and scores equal or below 2.39 as ‘poor’ or lacking fidelity.

Data analysis

Data was initially entered into an Excel spreadsheet and later exported onto an SPSS database. All analyses were conducted using IBM SPSS Statistics version 25 for Mac (IBM Corporation, 2017). Service variables were coded in the following way:

'Team' (1-12), 'Model of care' (1=OD, 2=CAU), 'Weekly meetings' (Yes/No), 'Weekly individual supervision' (Yes/No), and 'Monthly group supervision' (Yes/No) were entered as nominal variables. 'Number of employed staff members', 'Number of staff roles available', and 'Individual caseload' were coded as continuous variables. All CoMFideS items were coded as ordinal variables and then transformed into their respective sections. All CoMFideS sections were coded as continuous variables. Descriptive statistics and radar plots were used to characterise the participating sites.

Psychometric properties. The present study explored –albeit tentatively– the following psychometric properties of the CoMFideS: (1) response bias, (2) inter-rater reliability, (3) internal consistency, (4) face and content validity, and (5) construct validity. Other forms of measure validity were not possible given the scarcity of data. Further, convergent, divergent, and criterion (i.e. predictive) validity testing were not possible due to a lack of (1) validated OD fidelity measures, and (2) outcomes from the ODDESSI trial, respectively. Similarly, a factor analysis was not possible. Literature suggests that an *absolute* minimum sample size for factor analysis calculations should be *at least* 3 times the number of items (i.e. 32 items * 3 = 96 responders) (MacCallum, Widaman, Zhang, & Hong, 1999; Mundfrom, Shaw, & Ke, 2005). Given the lack of additional OD services in the country this was impossible to obtain, thus leaving factor analysis for later stages of the trial when repeated measures of CoMFideS have been completed.

Response bias. Response bias was explored using the range of scores generated by both respondents (i.e. managers and practitioners) from the participating sites ($n=24$). Score frequencies were obtained for each item on a respondent level. Items with the same score on more than 90% of respondents were considered to have inadequate

response spread. Likewise, given that the study was part of wider research trial requiring a certain standard of care, a somewhat tight grouping (i.e. negative skew) of scores was expected.

Reliability. In terms of inter-rater reliability, it was decided not to obtain Pearson's r or intra-class coefficients (ICC) given the sample size, and that respondents and raters were not fully crossed or nested. Neither of these tests can remove systematic coder deviations and can therefore underestimate the true reliability of ill-structured measurement designs (ISMDs) such as the one used for this study (Hallgren, 2012; Putka et al., 2008). The G estimation coefficient (Putka et al., 2008) was therefore chosen to make up for the limited data and as a less biased reliability estimator of ratings for both types of respondents ($n=48$). The G coefficient explicitly models the effects associated to specific individual raters and, as such, it allows to distinguish between rater main effects and ratee-rater interaction effects and residual error. A G coefficient above 0.7 was considered acceptable. Additionally, internal consistency reliability was assessed using Cronbach's alpha (Cronbach, 1951) using all pairs of interviews ($n=24$). According to Streiner (2003), a Cronbach's alpha above 0.7 is considered acceptable.

Validity. Face and content validity were assumed as adequate given the iterative feedback and input from experts, managers, and staff members. However, the "known groups" method (Portney & Watkins, 2000) of construct validity was used to strengthen this assumption. The "known groups" method suggests that an instrument should be able to discriminate between subjects who are known or expected to have an attribute (i.e. OD training) from those who are not (i.e. CAU). Given the small amount of team-level scores ($n=12$), descriptive statistics of the OD addendum were used to tentatively test this assumption.

Results

Response rates

All CAU and OD interviews were completed (100%) with no missing data. The response rate was 100% (12 of 12 pairs of interviews). Data was collected from 12 of the 12 possible teams.

Service characteristics

Only CAU teams were able to provide copies of their operational policies as OD teams were still in the process of developing their own; however, given the structure of the trial clusters, CAU policies were also taken to apply to OD teams. All teams had a clear understanding of the referral and eligibility criteria, maximum caseloads, and training and supervision arrangements. The average caseload per staff member was 25.8 service-users ($SD=7.36$, range=20–40) for the OD teams and 29.8 ($SD=8.50$, range=25–45) for CAU teams. The mean staff positions for OD teams was 9.5 ($SD=3.08$, range=5-13) and for CAU teams was 13.8 ($SD=3.49$, range=10-19). Psychiatrists, psychiatric nurses, clinical psychologists and psychotherapists were the most common professions and were all employed across teams ($n=6$). Occupational therapists were employed by 83% ($n=5$) of CAU teams, whereas only in 50% of OD teams. Only one CAU team (8%) employed advocates. Nurse assistants were employed by 25% of the teams ($n=3$) altogether (Table 2).

Preliminary psychometric properties of the CoMFideS measure

Response bias. Overall there seemed to be an even distribution of scores across the CoMFideS measure ($n=24$) with a mean variance of scores of 0.85 ($SD=0.55$, range=0.15-2.25). None of the items had more than 90% of responders receiving the same score. The only near exceptions were items ‘Supervision’ and ‘Flexibility of

Table 2. Service characteristics.

	Open Dialogue (n=6)		Care as usual (n=6)	
	\bar{X} (Range)		\bar{X} (Range)	
Employed staff (FTE and WTE)	9.50(5-13)		13.82(10-19)	
Caseload	\bar{X} (SD)		\bar{X} (SD)	
Team	220.83(120.68)		503.33(165.73)	
Individual	25.83(7.36)		29.83(8.50)	
	<i>n</i>	%	<i>n</i>	%
Service setup				
Integrated	5	83.3	0	0.0
Stand-alone	1	16.7	6	100.0
Staff roles				
Psychiatrists	6	100.0	6	100.0
Nurses	6	100.0	6	100.0
Nurse Assistants	2	33.3	1	16.7
Psychologists	6	100.0	6	100.0
Occupational Therapists	3	50.0	5	83.3
Social Workers	3	50.0	4	66.7
Support Workers	3	50.0	5	83.3
Peer Support Workers	6	100.0	1	16.7
Advocates/Volunteers	0	0.0	1	16.7
Weekly team meetings	6	100.0	6	100.0
Supervision arrangements				
Individual	5	83.0	6	100.0
Group	6	100.0	3	50.0

response’, both of which had a score of 4 for 83.3% and 87.5% respondents respectively. 20 items (63%) were negatively skewed, with only one the item ‘Flexibility of response’ being beyond the acceptable limits of ± 2 (Field, 2017). 23 items (72%) were leptokurtic, with only the item ‘Continuing professional development’ being beyond acceptable limits (kurtosis=2.13). For a detailed account of score distributions see Appendix K. In terms of differences between respondents, managers and practitioners had similar response ranges, with managers scoring marginally higher on all 5 sections (Table 3). On the 25-item CoMFideS, managers had a mean score of 3.17 (SD=0.45, range=2.52–3.88) whilst practitioners had a mean score of 3.05 (SD=0.40, range=2.52–3.56).

Table 3. Differences in CoMFideS scores between respondents.

Section	Managers		Practitioners	
	Mean (SD)	Range	Mean (SD)	Range
CoMFideS (25-item)	3.17(0.45)	2.52-3.56	3.05(0.40)	2.52-3.88
Service structure and culture	3.08(0.39)	2.25-3.63	2.93(0.39)	2.63-3.75
Access and engagement	3.24(0.54)	2.33-3.67	3.18(0.44)	2.33-4.00
Delivery of care	3.04(0.64)	2.00-4.00	2.96(0.61)	2.00-4.00
Community linkage and support	3.37(0.58)	1.00-3.71	3.20(0.53)	2.20-4.00
Open Dialogue addendum	2.48(1.23)	2.52-3.56	2.26(1.11)	1.00-4.00

Reliability analysis. Item-level calculations of the G estimate of reliability suggested a potentially good inter-rater reliability across the measure. 22 of the 32 items (68.7%) showed coefficients above 0.9 (Table 4). The only three exceptions were items ‘Providing information’, ‘Flexibility of response’, and ‘Coordination of care’ which had reliability coefficients of 0.69 and 0.65 respectively. Likewise, item on OD CPD had a coefficient of 0 given its null variance (rate variance=0.000, rater variance=0.000, estimated variance of the combination of rate*rater interaction and residual effects=2.298).

Internal consistency. Both the 25-item CoMFideS scale and the 7-item OD addendum suggested potentially good internal consistency, with Cronbach’s alpha coefficients of 0.90 and 0.95 respectively (Table 5). An item-level analysis was conducted to examine whether deleting any individual item would make important changes to the overall internal consistency of each scale. Results suggested little influence of any individual item on the total internal consistency of the 25-item CoMFideS scale (coefficient change ranging from -0.002 to 0.01) and the 7-item OD addendum (range=-0.020–0.016).

When analysed on a section level, all 5 sections appeared to have adequate

Table 4. Inter-rater reliability of the CoMFideS using the G estimate (n = 24)

Item	G(0.200, 2)
CoMFideS scale	0.992
SSC1. Service ethos and comprehensiveness	0.914
SSC2. Staff training	0.868
SSC3. Supervision	0.829
SSC4. Staff roles	0.918
SSC5. Staff capacity	0.897
SSC6. Routine outcome monitoring	0.952
SSC7. Safety	0.896
SSC8. Service-user involvement in co-production	0.944
AE1. Access to the service	0.927
AE2. Providing information	0.689
AE3. Prompt action	0.818
AE4. Identification of support systems	0.916
AE5. Flexibility of response	0.421
AE6. Assertive engagement	0.913
DC1. Continuity of care	0.896
DC2. Establishing clinical meetings	0.918
DC3. Collaborative decision making	0.950
DC4. Information sharing and communication	0.751
DC5. Service-user involvement in the delivery of care	0.829
DC6. Coordination of care	0.646
CLS1. Service linkage	0.884
CLS2. Community links (Practitioner level)	0.783
CLS3. Community links (Support system)	0.929
CLS4. Caregiver involvement and support	0.969
CLS5. Discharge and aftercare	0.760
Open Dialogue addendum	0.997
OD1. Transparency	0.929
OD2. Self-disclosure	0.970
OD3. Intervision frequency	0.990
OD4. Intervision content and structure	0.995
OD5. Team self-work	0.964
OD6. OD Training	0.995
OD7. OD Continued Professional Development	0.000

internal consistency (Table 5). Results suggested little influence of any individual item on the total internal consistency of their respective section (coefficient increases ranging from 0.02 to 0.04 across sections); however, some items showed very small item-total correlations (minimum value of 0.3; Field, 2017). Further, some items were found to

Table 5. Internal consistency of the CoMFideS.

CoMFideS section (<i>n</i> =24)	Internal consistency (Cronbach's alpha)
Service structure and culture	0.681
Access and engagement	0.677
Delivery of care	0.817
Community linkage and support	0.713
Open Dialogue addendum	0.954

negatively correlate with their sub-samples. For instance, on the ‘Service structure and culture’ section, items ‘Supervision’ had a negative item-total correlation of -0.01 as did ‘Training’ and ‘Staff roles’, with coefficient of -0.25 and -0.29 respectively. Also, on the ‘Access and engagement’ section, item ‘Flexibility of response’ had a negative item-total correlation of -0.04. All other item-total correlation coefficients were above 0.4.

Known groups validity. Although both OD and CAU teams scored similarly across the measure, OD teams unsurprisingly had higher scores in the Open Dialogue Addendum compared to CAU teams (Table 6). For the Open Dialogue Addendum, the mean score of the 6 OD teams was 3.44 (*SD*=0.36), whereas the 6 CAU teams had a mean score of 1.30 (*SD*=0.30).

*Table 6. Differences in CoMFideS mean scores between service models (*n*=12)*

	Open Dialogue (<i>n</i> =6)		Care as usual (<i>n</i> =6)	
	Mean (SD)	Range	Mean (SD)	Range
CoMFideS score	3.25(0.38)	2.78 - 3.72	2.97(0.35)	2.72 - 3.66
Service structure and culture	3.02(0.37)	2.56 - 3.44	2.99(0.35)	2.63 - 3.63
Access and engagement	3.26(0.40)	2.58 - 3.75	3.15(0.44)	2.58 - 3.83
Delivery of care	3.35(0.51)	2.67 - 4.00	2.65(0.48)	2.17 - 3.50
Community linkage and support	3.47(0.34)	3.10 - 3.90	3.10(0.44)	2.60 - 3.70
Open Dialogue addendum	3.44(0.36)	2.93 - 3.79	1.30(0.30)	1.00 - 1.86

Item scores. On an item level, 6 of the 25 CoMFideS items (24%) had mean scores equal or above 3.40 (‘very good’); 14 items (56%) had scores between 2.80 and

3.39 ('good'); two items (8%) had scores between 2.40 and 2.79 ('acceptable'); and three items (12%) had scores below 2.39 ('poor') (Table 7).

Table 7. Lowest (left) and highest (right) scoring items across teams.

Mean scores between 2.00 and 2.39 (Poor fidelity)		Mean score above 3.40 (Very good fidelity)	
SSC6. Routine outcome measurements	2.08	SSC2. Staff training	3.71
SSC8. Co-production	2.13	SSC3. Supervision	3.83
DC5. Service-user involvement in delivery of care	2.38	AE5. Flexibility of response	3.88
		AE6. Assertive engagement	3.58
		CLS2. Community links (Practitioner level)	3.50
		CLS4. Caregiver involvement and support	3.42
		OD3. Intervention frequency (n=6)	4.00
		OD4. Intervention content and structure (n=6)	3.75
		OD6. OD Training (n=6)	3.67
		OD7. OD Continued professional development (n=6)	3.83

Standard of care (CoMFideS score)

Overall, the mean CoMFideS total score (i.e. excluding the OD addendum) across all 12 teams was 3.11 ($SD=0.38$, range=2.72–3.72), possibly suggesting 'good' fidelity to standard NHS care (Figure 2). When analysed by model of care, the 6 OD

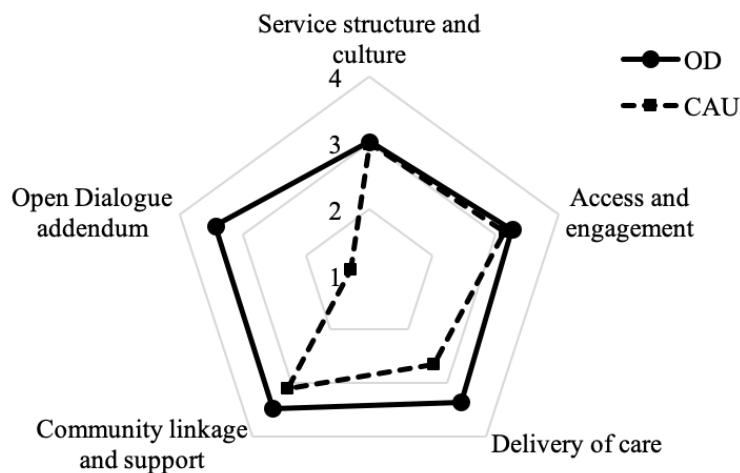


Figure 2. Comparison of mean CoMFideS section scores between Open Dialogue (OD) and Care as Usual (CAU).

teams had a mean CoMFideS total score of 3.25 ($SD=0.38$; range=2.78–3.72), whereas the 6 CAU teams had a mean CoMFideS total score of 2.97 ($SD=0.35$, range=2.72–3.66). Open Dialogue teams had higher scores on all sections compared to CAU teams.

Overall, OD teams scored higher on most items (Figure 3). CAU teams scored higher than OD teams in ‘co-production’ (mean=2.25, $SD=0.52$), ‘service capacity’

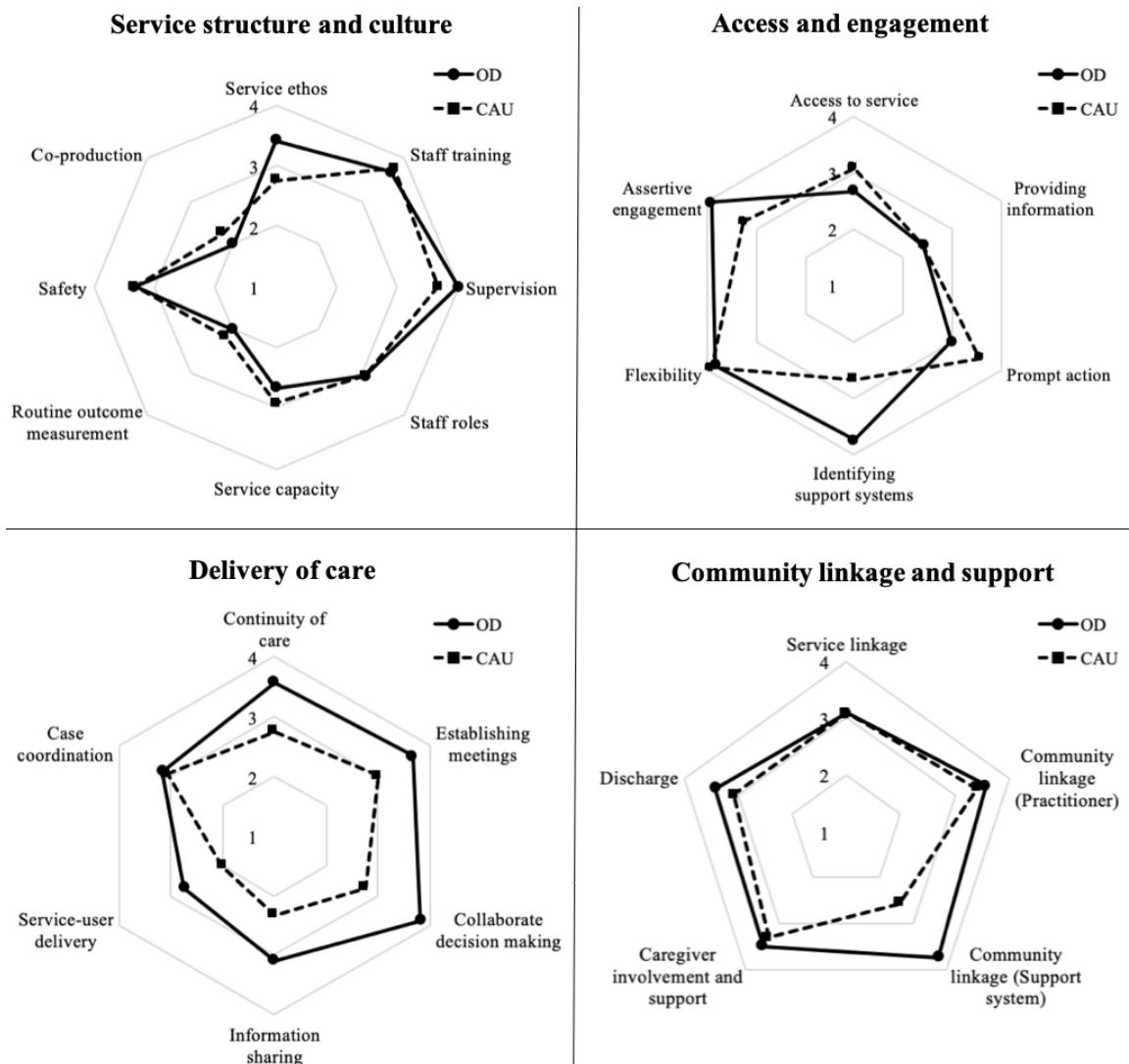


Figure 3. Comparison of mean CoMFideS item scores between Open Dialogue (OD) and care as usual (CAU).

(mean=2.92, $SD=0.49$) ‘routine outcome measurement’ (mean=2.17, $SD=0.26$), ‘access to the service’ (mean=3.08, $SD=0.66$), and ‘prompt action’ (mean=3.58, $SD=0.58$). For a detailed account of mean scores, standard deviations, and score ranges see Appendix L.

Open Dialogue fidelity

When focusing only on the 6 OD teams (i.e. the focus of this section), 3 of the 6 teams (50%) showed ‘very good’ fidelity, 2 teams (33%) were in the ‘good’ range, and one team (17%) demonstrated ‘acceptable’ fidelity. On an item level, 4 of the 7 items (57.1%) had mean scores equal or above 3.40 (‘very good’); two items (14.2%) had scores between 2.80 and 3.39 (‘good’); and one item (14.2%) had scores between 2.40 and 2.79 (‘acceptable’) (Figure 4).

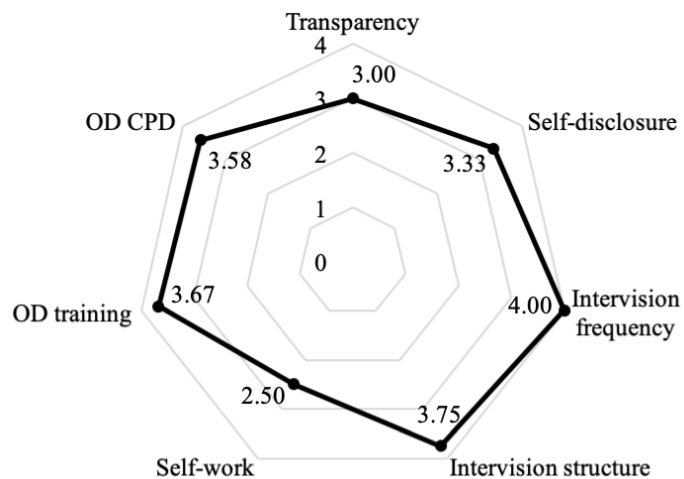


Figure 4. Mean scores of the Open Dialogue (OD) addendum (Open Dialogue teams). CPD=Continued professional development.

Differences in implementation across sites

All 12 teams had similar scores across the whole CoMFideS measure (Table 8). Team 6 had the lowest CoMFideS total score (mean=2.72, $SD=0.74$) which corresponds to an ‘acceptable’ fidelity score. Team 12 had the highest CoMFideS total score

(mean=3.72, *SD*=0.41) corresponding to a ‘very good’ fidelity score. Further, focusing only on OD teams, Team 9 had the lowest mean score on the Open Dialogue addendum (mean=2.93; *SD*=1.13), representing a ‘good’ fidelity score and, once again, Team 12 had the highest score (mean=3.79, *SD*=0.39) representing a ‘very good’ fidelity score.

Table 8. CoMFideS mean scores across teams.

	CoMFideS scale (25-item)	Service structure and culture	Access and engagement	Delivery of care	Community linkage and support	Open Dialogue addendum
Care as usual						
Team 1	2.84(0.73)	2.63(0.79)	2.58(0.74)	2.75(0.52)	3.60(0.42)	1.31(0.00)
Team 2	3.66(0.49)	3.63(0.58)	3.83(0.41)	3.50(0.55)	3.70(0.45)	2.06(1.03)
Team 3	2.84(0.66)	2.81(0.65)	3.25(0.52)	2.42(0.74)	2.90(0.55)	1.56(0.48)
Team 4	3.02(0.64)	3.00(0.65)	3.42(0.80)	2.75(0.42)	2.90(0.55)	1.38(0.39)
Team 5	2.74(0.84)	2.81(0.88)	2.92(1.02)	2.33(0.75)	2.90(0.74)	1.38(0.39)
Team 6	2.72(0.74)	3.06(0.68)	2.92(0.80)	2.17(0.41)	2.60(0.82)	1.44(0.19)
Open Dialogue						
Team 7	3.00(0.79)	2.69(1.03)	3.17(0.82)	3.08(0.58)	3.20(0.57)	3.07(0.84)
Team 8	3.42(0.57)	3.06(0.73)	3.50(0.45)	3.67(0.52)	3.60(0.22)	3.71(0.39)
Team 9	2.78(0.82)	2.56(1.02)	2.58(0.86)	3.00(0.63)	3.10(0.65)	2.93(1.13)
Team 10	3.58(0.47)	3.44(0.32)	3.42(0.74)	3.67(0.41)	3.90(0.22)	3.64(0.48)
Team 11	2.98(0.96)	2.94(1.18)	3.17(0.93)	2.67(0.98)	3.20(0.76)	3.50(0.41)
Team 12	3.72(0.41)	3.44(0.50)	3.75(0.42)	4.00(0.00)	3.80(0.27)	3.79(0.39)

Discussion

The Community Mental Health Fidelity Scale

These preliminary findings suggest that CoMFideS could be a robust measure of programme fidelity for crisis and continued community care teams. This in part draws on its roots on the CYPRESS measure, which was found to be effective in assessing MST fidelity (Gaffney, 2012). As expected, most items were negatively skewed and leptokurtic, given that all teams were expected to be at least on an ‘acceptable’ level of fidelity. Considering this study was the first attempt at defining (and refining) the

content and scope of the measure, it is interesting to note that (a) only two items suggested possible response bias, and (b) only two items were beyond acceptable distribution limits. Nevertheless, as previously stated, this is likely due to the fact that this study was part of a wider research trial where one might expect a somewhat tight grouping of scores.

In terms of reliability, inter-rater reliability looks potentially promising. Although two items were marginally below acceptable ranges, it is possible that this was a consequence of unclear behavioural anchors. Interestingly, ‘providing information’ and ‘coordination of care’ received mixed feedback from experts. Developers argued that providing information about the service to clients and referrers helps streamline access to the service; however, SP felt that these two features were too similar to tease them apart during interviews. Similarly, SP and RR considered that coordination of care was a key component of crisis and continued community care, while RR and MA –with this same argument in mind– suggested its removal arguing it would turn out to be redundant. With regards to ‘flexibility of response’, this may have been due to the lack of clarity in the definition, which made it difficult for raters to reach a consensus in scores. Future versions of the manual could include a clearer definition of the item and a more specific behavioural anchors.

On a parallel note, the lack of variance on the ‘OD CPD’ item was unexpected. MH and RR expected Open Dialogue teams to have varying degrees of CPD participation, based on how much staff were encouraged (and supported) to attend. Feedback from interviews seemed to suggest that, in fact, all staff members were *encouraged* to attend the annual OD CPD events but only a few staff members actually attended, primarily due to workload pressures. With this in mind, a possible solution

would be to either reconsider the item as a dichotomous variable (i.e. whether team representatives attend or not) or –ideally– to redefine the anchors so that they focus on actual attendance rates.

Considering validity, CoMFideS appears to have adequate face and content validity and the ODDESSI team considered it feasible for the full trial. The iterative item refinement process, as well as the discussions with international experts in the field (including the developer of Open Dialogue) were central to developing items that would fit both models of care whilst also being sensitive enough to distinguish between them. Our ‘known groups’ method for construct validity was an attempt at supporting this assumption. However, with such a small sample size, caution must be applied moving forward.

Defining ‘good’ standard of care

On the question of standard of care, results might suggest that all teams demonstrated a ‘good’ standard of care; however, there is still some uncertainty about what this term really implies. As per the trial protocol, most OD teams (with the exception of Team 12, which was an independent service prior to the trial) emerged from CAU teams; nonetheless, there was a varying degree of experience, staffing, and capacity across teams. We expected the measure would be able to detect these intricacies by classifying teams along different levels of fidelity.

Our four-tier cut-off approach was chosen as a plausible –yet arbitrary– solution based on existing literature on fidelity measurement. Although it proved useful in determining whether participating teams were ready for inclusion in the trial (i.e. ‘acceptable’ fidelity) it did not seem to detect the expected variations in fidelity. Perhaps stricter score ranges could identify these nuances in future versions of the CoMFideS.

For instance, although a 5% increase on each cut-off range would have maintained all 12 teams in the ‘good fidelity’ range (mean score=3.11) there would have been differences on a model-level; whilst OD teams would have remained as ‘good’ (mean score=3.25) the CAU teams would have dropped to ‘acceptable’ (mean score=2.97). Further, the distribution of scores would have shifted considerably, with only two teams (OD=1, CAU=1) scoring as ‘very good’, 4 teams (OD=3, CAU=1) as ‘good’, and 6 teams (OD=2, CAU=4) as ‘adequate’. Although reasonable, this cut-off structure would indeed be just as arbitrary as the one presented here; therefore, a decision would best be made after collecting additional data from the trial.

Differences in implementation

Whether there are any consistent differences in implementation between each model’s teams is still unclear but will be monitored over time in the trial. The scarcity of data did not allow to identify specific patterns of implementation across sites. However, based on the integration of findings, it might still be possible to make some preliminary remarks:

Service structure and culture. There did not seem to be a particularly consistent view of the model of care in CAU teams. Both managers and practitioners provided different accounts about the way their team worked and their approach to crisis and continued care. This was not the case for OD teams, where all staff members had a more consistent view of the Open Dialogue approach and how it was being provided in their respective services. This may have been because of the multiple theoretical models and trainings involved in CAU teams and, perhaps, because of broader role responsibilities. Whichever the case, some staff seemed to have more directive and resource-oriented perspectives about their work and others a more person-centred stance.

Nevertheless, both teams found their training, supervision, and safety protocols (both for themselves and their service users) were appropriate which was reassuring for the trial and expected, given nation-wide standards of practice. On another note, although all teams considered that the staffing and distribution of professional roles inside their teams was adequate, OD teams described being pressured by their caseloads. It is possible that this is merely a reflection of the novelty of the approach (which could be solved with an increased number of staff) or it might be a reflection of how resource-intensive this model of care might be (Razzaque & Wood, 2015). Finally, there was a clear lack of attention across the board to outcomes measurement (including service user feedback), and only few accounts of team efforts to include service users in the development and planning of their teams (i.e. co-production). This was surprising given the current national efforts in encouraging routine outcome measurement and service-user involvement strategies for service development (Mockford, Staniszewska, Griffiths, & Herron-Marx, 2011; Richardson et al., 2019).

Access and Engagement. Overall, CAU staff were clearer about their referral pathways and inclusion/exclusion criteria compared to OD staff. This was expected as OD teams are yet to consolidate this in their operational policies as part of WP1. It might be possible that the lack of clarification of exclusion criteria –alongside the seemingly small efforts in informing referrers about their service remit– explains why OD teams feel understaffed and overwhelmed. Notwithstanding the access concerns, both CAU and OD teams seemed to be equally successful in providing immediate support to service users in crisis (at least by getting in touch with them within the first 24 hours), in making multiple efforts to engage their service users in treatment, and in providing multiple types of support depending on their individual needs. However, OD services

were considerably more proactive in involving the service users' support networks, which clearly reflected the 'social networks perspective' of the model (Seikkula et al., 1995).

Delivery of care. Psychological continuity (i.e. continuity of care) was indeed a much reinforced aspect of care for OD teams. Although CAU services did in fact recognise it as important to support their service users, it was not a priority or requirement for these services. Unexpectedly, both models of care had a very similar approach to establishing clinical meetings, being both seemingly just as flexible and eager to accommodate to the service users' requests for scheduling, location, and attendees. This flexibility though was not as apparent when related to decision-making processes; CAU teams were relatively more directive than their OD counterparts. Where both teams did somewhat falter, was in relation to sharing information and making communication transparent with everyone involved, which was unexpected from OD teams as this item reflects an important aspect of service-level 'transparency' (Olson et al., 2014; Razzaque & Stockman, 2016). Perhaps the workload makes it harder for some practitioners to keep track of communications carried out with other service providers. These matters aside, both models of care seemed equally conscientious in their coordination of care.

Community linkage and support. Results suggest that, although both teams are equally good in involving the whole community in the treatment and discharge process, OD teams seemed more successful in displacing the locus of agency towards the service users' own support systems. In other words, OD teams seem to make a bigger effort in empowering their service users by means of empowering the network around them (e.g. finding ways of encouraging friends or family to take the service users for GP

appointments rather than assigning a care worker to do it for them). This is an important finding, especially considering the resource constraints of community mental health teams and the global calls for sustainability of treatment outcomes (Patel et al., 2018).

Open Dialogue in the NHS

Our findings seem to suggest that crisis and continued community care services, once reorganised into the OD model of care, might be able to deliver OD with sufficient fidelity to the model *whilst also* operating effectively within standard NHS care principles. Not only did OD teams seem to work with sufficient fidelity to both principles of care; they also appear to be capable of doing so more effectively than CAU teams. It is important to bear in mind that this difference might simply be a matter of sampling bias. However, what was striking was that the ‘Delivery of care’ section scored higher in OD teams. It is unclear whether something about OD practice makes it easier for these teams to observe recommended NHS practice principles or whether additional factors, such as a greater availability of resources for training and supervision, could explain this trend.

On another note, the items where OD teams seemed to score consistently higher focused on key components of network working, specifically (1) identification of service-user’s support systems, (2) the process of establishment of clinical meetings, (3) engagement of service users’ own support systems, (4) collaborative decision making, and (5) continuity of care. These are all features theorized as essential to crisis and continued community care but are thought to be crucial to best dialogic practice (Bergström et al., 2018; Olson et al., 2014; Seikkula & Arnkil, 2006). A factor analysis later in the trial could help better understand this phenomenon.

Turning now to whether it is possible to distinguish OD teams from standard crisis and continued community care teams, results suggest that –as far as the CoMFideS could assess– it might be possible to do so. OD teams scored significantly higher in the Open Dialogue addendum compared to CAU. What stood out, nonetheless, was to find that team self-work was the lowest scoring item across OD teams. Self-work practices are theorised to be central to dialogical practice (Razzaque & Stockman, 2016). It is plausible that either (a) existing OD teams are not currently as focused on self-work as they should be, or (b) that self-work is not as exclusive to Open Dialogue in the United Kingdom as expected. Additional data is necessary to establish whether this is a trend; however, this finding flagged the need for further efforts in the operationalisation of this variable and –potentially– to enhanced self-work efforts in OD teams.

Overall, according to the Open Dialogue addendum, training efforts for the trial seem to have accomplished a consistent ethos across OD teams. It was interesting that managers and practitioners were equally cautious when describing how their services enacted said principles, usually suggesting that some of these principles (e.g. transparency and self-disclosure) were dependent on the service’s experience in dialogic practice. It would be interesting to test whether this is the case once repeated measures have been collected and alongside results from the adherence study by ML.

Strengths and limitations

Thus far this thesis has argued that CoMFideS is potentially a feasible and reliable measure for use in the ODDESSI programme. Its development process and results from the present study suggest multiple strengths but also highlighted some limitations on the measure development process, the study design, and the measure itself.

Perhaps one of the main strengths of the study is in relation to the measure development process. Having the opportunity to discuss and revise the measure with the help of experts in the field allowed for a rich discussion about the theoretical ‘critical components’ of the Open Dialogue approach. This was a challenging task given the striking similarities between approaches. A modified Delphi approach to expert feedback (Dalkey & Helmer, 1963) would have been ideal to provide more structure to the measure development process; however, time pressure to pilot and implement the measure for the trial did not allow for this to happen. Nevertheless, the iterative nature of the development procedure and the confirmatory nature of the ‘known groups’ method were considered as enough evidence of the measure’s preliminary validity. It would still be reasonable –once additional data has been collected and the measure further revised– to carry out a survey for the wider Open Dialogue community to assess the acceptability of the items as meaningful markers of programme fidelity.

Second, in spite of the small number of trial sites, all data was successfully collected; this was due to the joint efforts of trial coordinators and research assistants who helped identify staff and arrange the interviews. A larger sample would have certainly been preferable, as it would have allowed for a more robust data analytic methodology (e.g. factor analysis) and more meaningful results; this was simply not available for this study.

Third, given that participants from each site (i.e. managers and practitioners) were selected based on availability, there may have been some degree of selection bias; however, by triangulating information from different sources (i.e. a pair of managers, a pair of practitioners, and a pair of raters) and by allowing raters to reach consensus, we expected to control for biased responses from individual interviewees.

As previously discussed, another limitation of this study was that raters were not fully crossed or nested given the difficulties in matching respondent and rater availability. This was further complicated with the potential for rater bias as the main author was the only consistent rater across measurements. This limitation was addressed in two ways: first, the G estimator –although unconventional– seemed a robust solution to this as it takes into account rater assortment and systematic rater deviations; and second, by recording interviews it is possible to subsequently assess whether independent raters reach similar scores. Perhaps future validation studies could try a more structured rating procedure in order to be able to obtain the more widely accepted intra-class coefficients.

Finally, although face and content validity were acceptable, some items do need to be revised to capture the more discrete features of both OD and CAU. The decision to not exclude any items from the measure thus far was because, at this stage, it is unclear whether their score distributions and psychometric properties were due to sample size or the items themselves. Nevertheless, the CoMFideS manual included as Appendix J already includes some changes that could help improve the measure. Additionally, it would be ideal to assess convergent validity once more data has been collected (which would also allow confirmatory factor analyses) and predictive validity, once outcomes become available.

Clinical implications

As stated in the introduction, in order to adapt one intervention from one context to another, it is important to understand its ‘essential components’. Therefore, perhaps the main clinical implication of this study is that it helped better understand which structural features of Open Dialogue applied to the NHS. Discussions with experts and

the key domains identified on Chapter 1 allowed to translate theoretical principles into structural aspects of NHS service delivery. However, not all of these features turned out to be exclusive to the approach, and some of them were not structurally identifiable. For instance, the principles of ‘person-centredness’, ‘immediate help’, and ‘responsibility’ – all important for any form of crisis care– did not seem key markers of the OD approach as suggested by the literature (Olson et al., 2014; Seikkula et al., 1995; Ziedonis et al., 2015). However, what did seem to be highlighted by this study was the model’s flexibility (in terms of meeting arrangements and treatment provision), overall transparency (e.g. “nothing about us without us” philosophy; Hopfenbeck, 2018) and an open stance towards self-disclosure in clinical practice.

Future research

Some features of dialogic practice did not lend themselves as easily for assessment. For instance, it is unclear whether “using OD as a mindful way of being in clinical and non-clinical work” can be objectively assessed (Ziedonis et al., 2015). Our items ‘service ethos and comprehensiveness’ and ‘team self-work’ aimed to explore this, but will need further work. Other features, such as most of the 12 key elements of dialogic practice (Olson et al. 2014), and the principle of ‘dialogue and polyphony’ (even though it was somewhat addressed as ‘transparency’ and ‘self-disclosure’) were deemed more relevant to clinician level adherence rather than to organizations themselves (see ML’s dissertation for the adherence measure). It would thus be interesting to see how both measures of treatment integrity (i.e. fidelity and adherence) fit together, alongside the trial’s overall process evaluation results.

In spite of the encouraging findings from this study, a full validation study

with a larger sample size would be necessary before using CoMFideS in other formats. Some emerging questions would be, for instance, whether CoMFideS can provide just as reliable results with only one interviewee. Also, it is unclear whether other data collection methods, such as telephone interviews would be as reliable. Finally, and perhaps most importantly, continued validation studies would improve the quality of the measure and help establish whether CoMFideS could be used in a wider range of community mental health services.

Conclusion

The present study developed, piloted and tested a programme fidelity measure for its use on the ODDESSI programme trials. The Community Mental health Fidelity Scale (CoMFideS) captures both standard NHS crisis care principles and Open Dialogue practice principles. The measure development process made use of recognised and recommended methods and used multiple raters, multiple sources, and multiple contexts to assess its properties. Preliminary psychometric results were encouraging, suggesting that CoMFideS might be amenable for its wider use in other community mental health settings where additional models of care are being trialled. Results also suggest that CoMFideS may be able to not only establish (a) the extent to which teams deliver their respective models according to established protocols, but also (b) the degree of differentiation between two very similar approaches to crisis and recovery. Furthermore, CoMFideS seems to be feasible for large scale programme evaluations, as it requires less than an hour to complete.

References

Bakhtin, M. (1981). *The dialogic imagination: Four essays*. Austin, TX: University of Texas.

- Bateson, G., Jackson, D., Haley, J., & Weakland, J. (1956). Toward a theory of schizophrenia. In C. Sluzki & D. Ransom (Eds.). *Double bind: The foundation of the communicational approach to the family*. New York, NY: Grune & Stratton.
- Bergström, T., Seikkula, J., Alakare, B., Mäki, P., Kögäs-Saviaro, P., Taskila, J., Tolvanen, A. & Aaltonen, J. (2018). The family-oriented open dialogue approach in the treatment of first-episode psychosis: Nineteen-year outcomes. *Psychiatry Research*, 270, 168-175. DOI: 10.1016/j.psychres.2018.09.039
- Bond, G., Becker, D., Drake, R., & Vogler, K. (1997). A fidelity scale for the individual placement and support model of supported employment. *Rehabilitation Counseling Bulletin*, 40, 265-284.
- Bond, G., Campbell, K., Evans, L., Gervey, R., Pascaris, A., Tice, S., Del Bene, D., & Revell, G. (2002). A scale to measure quality of supported employment for persons with severe mental illness. *Journal of Vocational Rehabilitation*, 17(4), 239-250
- Bond, G., Williams, J., Evans, L., Salyers, M., Kim, H., Sharpe, H., & Leff, H. (2000). *Psychiatric rehabilitation fidelity toolkit*. Cambridge, MA: Human Services Research Institute.
- Borelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, 71(S1), S52-S63. DOI: 10.1111/j.1752-7325.2011.00233.x
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9 (3), 458-467.

- Donabedian, A. (1988). The quality of care: how can it be assessed?. *Journal of the American Medical Association*, 260(12), 1743-1748.
- Essock, S., Nossel, I., McNamara, K., Bennet, M., Buchanan, R., Kreyenbuhl, J., Mendon, S., Goldman, H., & Dixon, L. (2015). Practical monitoring of treatment fidelity: Examples from a team-based intervention for people with early psychosis. *Psychiatric Services*, 66(7), DOI: 10.1176/appi.ps.201400531
- Field, A. (2017). *Discovering statistics using SPSS (5th Ed)*. London: SAGE.
- Gaffney, C. (2012). *Complex interventions for children and young people: exploring service delivery frameworks and characterising interventions* (Doctoral Dissertation, University College London). Retrieved from <http://discovery.ucl.ac.uk/1367068/1/C.GAFFNEY%20VOLUME%201.pdf>
- Gearing, R., El-Bassel, N., Ghesquierre, A., Baldwyn, S., Gillies, J., & Ngeow, E. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, 31, 79-88. DOI: 10.1016/j.cpr.2010.09.007
- Giacco, D., McCabe, R., Kallert, T., Hansson, L., Fiorillo, A., & Priebe, S. (2012). *Friends and symptom dimensions in patients with psychosis: a pooled analysis*. *PLoS One*, 7(11), e50119.
- Gillard, S., Holley, J., Gibson, S., Larsen, J., Lucock, M., Oborn, E., Rinaldi, M., & Stamou, E. (2015). Introducing new peer worker roles into mental health services in England: Comparative case study research across a range of organisational contexts. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(6), pp. 682-694.

- Gonzalez Castro, F., Barrera, M., & Martinez, C. (2004). The Cultural Adaptation of Prevention Interventions: Resolving Tensions Between Fidelity and Fit. *Prevention Science*, 5(1), 41-45.
- Hallgren, K. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23. DOI: 10.20982/tqmp.08.1.p023
- Hartman, D., & De Coursey, J. (2015). Family Therapy in the Real World: Dialogical Practice in a Regional Australian Public Mental Health Service. *Australian and New Zealand Journal of Family Therapy*, 36(1), 88-101.
- Holmbeck, G., & Devine, K. (2009). Editorial: An author's checklist for measure development and validation manuscripts. *Journal of Pediatric Psychology*, 34(7), 691-696. DOI: 10.1093/jpepsy/jsp046
- Holmesland, A., Seikkula, J., & Hopfenbeck, M. (2014). Inter-agency work in Open Dialogue: the significance of listening and authenticity. *Journal of Interprofessional care*, 28(5), 433-439.
- Hopfenbeck, M. (2015). Peer-supported Open Dialogue. *Context*, 138, 29-31.
- Hopfenbeck, M. (2018). Intervision: Supervision for Peer-supported Open Dialogue Teams. *Unpublished manuscript*.
- IBM Corporation (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM.
- Jacobs, R., & Barrenho, E. (2011). Impact of crisis resolution and home treatment teams on psychiatric admissions in England. *The British Journal of Psychiatry*, 199(1), pp. 71-76.
- Jethwa, K., Galappathie, N., & Hewson, P. (2007). Effects of a crisis resolution and

home treatment team on in-patient admissions. *Psychiatric Bulletin*, 31(5), 170-172.

Johnson, S. (2013). Crisis resolution and home treatment teams: an evolving model. *Advances in Psychiatric Treatment*, 19 (2), 115-123.

Johnson, S., Nolan, F., Hoult, J., White, I., Bebbington, P., Sandor, A., McKenzie, N., Patel, S., & Pilling, S. (2005a). Outcomes of crises before and after introduction of a crisis resolution team. *The British Journal of Psychiatry*, 187(1), pp. 68-75.

Johnson, S., Nolan, F., Pilling, S., Sandor, J., Hoult, J., McKenzie, N., White, I., Thompson, M., & Bebbington, R. (2005b). Randomised controlled trial of acute mental health care by a crisis resolution team: the north Islington crisis study. *BMJ*, 331(7517), 599. DOI: 10.1136/bmj.38519.678148.8F

Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–6.

Lakeman, R. (2014). The Finnish open dialogue approach to crisis intervention in psychosis: A review. *Psychotherapy in Australia*, 20(3), 28.

Lloyd-Evans, B., Bond, G., Ruud, T., Ivanecka, A., Gray, R., Osborn, D., Nolan, F., Henderson, C., Mason, O., Goater, N., Kelly, K., Ambler, G., Morant, N., ONyett, S., Lamb, D., Fahmy, S., Brown, E., Paterson, B., Sweeney A., Hindle, D., Fullarton, K., Frerichs, J., & Johnson, S. (2016). Development of a measure of model fidelity for mental health Crisis Resolution Teams. *BMC psychiatry*, 16(1), 427. DOI: 10.1186/s12888-016-1139-4

Lloyd-Evans, B., Christoforou, M., Osborn, D., Ambler, G., Marston, L., Lamb, D., Mason, O., Morant, N., Sullivan, S., Henderson, C., Hunter, R., Pilling, S., Nolan, F., Gray, R., Weaver, T., Kelly, K., Goater, N., Milton, A., Johnston, E., Fullarton, E., Lean, M., Paterson, B., Piotrowski, J., Davidson, M., Forsyth, R.,

- Mosse, L., Leverton, M., O'Hanlon, P., Mundy, E., Mundy, T., Brown, E., Fahmy, S., Burgess, E., Churchard, A., Wheeler, C., Istead, H., Hindle, D., & Johnson, S. (2019). Crisis resolution teams for people experiencing mental health crises: the CORE mixed-methods research programme including two RCTs. *Programme Grants for Applied Research*, 7(1), 1-102. DOI: 10.3310/pgfar07010
- Lloyd-Evans, B., Mayo-Wilson, E., Harrison, B., Istead, H., Brown, E., Pilling, S., Johnson, S., & Kendall, T. (2014). A systematic review and meta-analysis of randomised controlled trials of peer support for people with severe mental illness. *BMC Psychiatry*, 14(39), 14-39. DOI: 10.1186/1471-244X-14-39
- MacCallum, R., Widaman, K., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84. DOI: 10.1002/anzf.1094
- Mikes-Liu, K. (2015). Is it Possible to be a Bit Dialogical? Exploring how a dialogical perspective might contribute to a psychiatrist's practice in a child and adolescent mental health setting. *Australian and New Zealand Journal of Family Therapy*, 36(1), 122-139.
- Mockford, C., Staniszewska, S., Griffiths, F., & Herron-Marx, S. (2011). The impact of patient and public involvement on UK NHS health care: a systematic review. *International Journal for Quality in Health Care*, 24(1), 28-38.
- Morton, C., & Norman-Nott, A. (2019). The interface with community services. In A. Barrera, C. Attard and R. Chaplin (Eds.) *The Oxford textbook of inpatient psychiatry*. England: Oxford University Press.
- Mundfrom, D., Shaw, D., & Ke, T. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2), 159-168. DOI: 10.1207/s15327574ijt0502_4

- NHS Confederation (2016). *Is mental health crisis in crisis?* (Issue brief 290). Retrieved from <https://www.nhsconfed.org//media/Confederation/Files/Publications/Documents/Is-mental-health-crisis-care-in-crisis-MHN-Briefing-290.pdf>
- Olson, M., Seikkula, J., & Ziedonis, D. (2014). *The Key Elements of Dialogic Practice in Open Dialogue: Fidelity Criteria*. The University of Massachusetts Medical School. Worcester, MA.
- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P., Cooper, J., Eaton, J., Herrman, H., Herzallah, M., Huang, Y., Jordans, M., Kleinman, A., Medina-Mora, M., Morgan, E., Niaz, U., Omigbodun, O., Prince, M., Rahman, A., Saraceno, B., Sarkar, B., De Silva, M., Singh, I., Stein, D., Sunkel, C., & Unützer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553-1598.
- Pilling, S. (2016). *Open Dialogue: Development and Evaluation of a Social Network Intervention for Severe Mental Illness (ODDESSI)*. Grant application for the National Institute for Health Research. London, England.
- Pilling, S., Butler, S., Gaffney, C., & Fonagy, P. (2012). CYPRESS measure (Children and Young People-Resources, Evaluation and Systems Schedule). Unpublished manuscript, University College London, UK.
- Pocobello, R., & Salamina, G. (September, 2015). *Introducing Open Dialogue in Italy: The role of participatory evaluation*. In «LISTEN TO ME!» -HUMANIZING HUMAN PRACTICES- Third International Conference on Dialogical Practices. Norway: Kristiansand. Retrieved from <https://goo.gl/NAVp4e>

- Portney, L., & Watkins, M. (2000). *Foundations of clinical research: applications to practice*. Upper Saddle River, NJ: Prentice Hall Health.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Giffey, R., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65-76.
- Putka, D., Le, H., McCloy, R., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology*, 93(5), 959. DOI: 10.1037/0021-9010.93.5.959
- Rasinkangas, A., & Lehtinen, V. (2003). Open dialogue approach: Treatment principles and preliminary results of a two-year follow-up on first episode schizophrenia. *Ethical Human Sciences and Services*, 5(3).
- Razzaque, R., & Stockmann, T. (2016). An introduction to peer-supported open dialogue in mental healthcare. *BJPsych Advances*, 22(5), 348-356.
- Razzaque, R., & Wood, L. (2015). Open Dialogue and its relevance to the NHS: Opinions of NHS Staff and Service Users. *Community Mental Health Journal*, 51(8), 931-938. DOI: 10.1007/s10597-015-9849-5
- Richardson, E., Walshe, K., Boyd, A., Roberts, J., Wenzel, L., Robertson, R., & Smithson, R. (2019). User involvement in regulation: A qualitative study of service user involvement in Care Quality Commission inspections of health and social care providers in England. *Health Expectations*, 22(2), 245-253.
- Santacroce, S., Maccarelli, L., & Grey, M. (2004). Intervention fidelity. *Nursing Research*, 52(1), 63-66.
- Schoenwald, S., Garland, A., Chapman, J., Frazier, S., Sheidow, A., & Southam-Gerow,

- M. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(1), 32-43. DOI: 10.1007/s10488-010-0321-0
- Seikkula, J., Aaltonen, A., Alakare, B., Haarakangas, K., Keränen, J., & Lehtinen, K. (2006). Five-year experience of first-episode nonaffective psychosis in open-dialogue approach: Treatment principles, follow-up outcomes, and two case studies. *Psychotherapy Research*, 16(2), 214-228. DOI: 10.1080/10503300500268490
- Seikkula, J., Aaltonen, A., Alakare, B., Haarakangas, K., Keränen, J., Sutela, M. (1995). Treating psychosis by means of open dialogue. In S. Friedman, *The reflecting team in action: collaborative practice in family therapy*. New York, NY: The Guilford Press.
- Seikkula, J., Alakare, B., & Aaltonen, A. (2001a). Open dialogue in psychosis I: An introduction and case illustration. *Journal of Constructivist Psychology*, 14(4), 247-265.
- Seikkula, J., Alakare, B., & Aaltonen, A. (2001b). Open dialogue in psychosis II: A comparison of good and poor outcome cases. *Journal of Constructivist Psychology*, 14(4), 267-284.
- Seikkula, J., Alakare, B., & Aaltonen, A. (2011). The comprehensive open-dialogue approach in Western Lapland: II. Long-term stability of acute psychosis outcomes in advanced community care. *Psychosis*, 3(3), 192-204.
- Seikkula, J. & Arnkil, T. (2006). *Dialogical meetings in social networks*. London: Karnak.

- Seikkula, J., & Olson, M. (2003). The open dialogue approach to acute psychosis: Its poetics and micropolitics. *Family process*, 42(3), 403-418.
- Seikkula, J., & Trimble, D. (2005). Healing elements of therapeutic conversation: Dialogue as an embodiment of love. *Family process*, 44(4), 461-475.
- Streiner, D. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99–103. DOI: 10.1207/S15327752JPA8001_18
- Teague, G., Mueser, K., & Rapp, C. (2012). Advances in fidelity measurement for mental health services research. *Psychiatric Services*, 63(8), 765-771. DOI: 10.1176/appi.ps.201100430
- The Commission on Acute Adult Psychiatric Care (2015) *Old Problems, New Solutions: Improving Acute Psychiatric Care for Adults in England*. Royal College of Psychiatrists. London: England.
- The Kings Fund (2016). *Deficits in the NHS 2016* (Briefing). Retrieved from https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Deficits_in_the_NHS_Kings_Fund_July_2016_1.pdf
- Weisman, G. (1989). Crisis intervention. In A. Bellack, *A Clinical Guide for the Treatment of Schizophrenia*. Boston, MA: Springer.
- Wheeler, C., Lloyd-Evans, B., Churchard, A., Fitzgerald, C., Fullarton, K., Mosse, L., Paterson, B., Galli Zugaro, C., & Johnson, S. (2015). Implementation of the Crisis Resolution Team model in adult mental health settings: a systematic review. *BMC Psychiatry*, 15(74), 1-14. DOI: 10.1186/s12888-015-0441-x
- Ziedonis, D., Olson, M., & Seikkula, J. (2015). *10 Organizational Criteria for Open Dialogue*. The University of Massachusetts Medical School. Worcester, MA.

Part 3: Critical Appraisal

Overview

This final chapter will focus on some final thoughts about the research project as a whole. I begin by discussing some methodological considerations of this research. Then, drawing from entries in my research log, I reflect on the research process and provide some context on how the project took shape. Following this, I share some insights about the Open Dialogue model and the potential role for anthropological research methods in better understanding the approach. Finally, I discuss the impact of my stance as a scientist-practitioner on this research, and the overall impact the project had on my research skills and clinical practice.

Methodological considerations

As I reflected on the results, I came to think that this project may have generated more questions than it answered. Firstly, although the systematic review offered a new typology for programme fidelity assessment and highlighted some useful domains to include on our Community Mental health Fidelity Scale (CoMFideS), it is still unclear what relationship these domains have with standards of care. Secondly, although these domains helped develop a measure that brought about a new understanding of crisis and continued community care teams in the NHS, it is unclear how our results may relate to adherence and the wider ODDESSI process evaluation. It remains for future studies to explore (1) whether a factor analysis confirms the structure of the suggested dimensions, (2) the relationship of these variables to actual outcomes from the trial, (3) the appropriateness of constructs (e.g. Delphi method), and (4) whether any response patterns emerge depending on the model of care.

Some questions about the role of language and culture in the development of fidelity measures still remain unanswered. ‘Severe mental health difficulties’, ‘quality of

care’, and ideas about which aspects of an intervention are most important (e.g. behaviours, skills, acumen, structure, technique, processes, etc.) are, to an extent, culturally-bound (Patel et al., 2018). This presented us with the complex task of defining a wide range of variables to allow for an empirical examination of what we considered to constitute –based on literature, expertise, and existing measures– a ‘good’ standard of crisis and continued community care in the NHS. Although helpful, and seemingly fruitful, this made me reflect on the risk of reifying concepts that might not necessarily be relevant, as it may have happened with our items of ‘providing information’ or ‘coordination of care’. This is not to say that these items are non-existent nor that they are unrelated to the models of care; however, they serve as an example of the impact that personal biases or professional assumptions may have on research, and the impact these may have on outcomes. It is yet to be established, by means of further exploration of the properties of CoMFideS, whether the items included (and their operationalisation) are in fact representative of the models we aimed to evaluate.

On that same note and echoing the results of our systematic review, this study was unable to deliver reliable psychometric data for the CoMFideS. Given its piloting nature and the lack of additional trial sites, this task was simply not possible. This makes me think that perhaps this is the case for many other –potentially useful and trustworthy– existing fidelity measures (such as CYPRESS or the TMACT) that, for some reason, did not reach full validation (Gaffney, 2012; Monroe-De Vita et al., 2011). I take this as a learning point about the importance of following through with the refinement and validation process of measures, given their potential usefulness for the field and for the interventions they evaluate.

A recent special issue of *The Lancet* (Patel et al., 2018) highlighted the

importance of enhancing efforts to expand the evidence-base for mental health interventions that target the social networks of people experiencing severe mental health difficulties. The development of the CoMFideS was an attempt at addressing this need. Results suggest that CoMFideS may have potential for helping us reliably study aspects of community mental health care. Further, given its similarity to CYPRESS (Gaffney, 2012), CoMFideS might capture elements that support a wide range of social networks interventions beyond the context of severe mental health difficulties. In this project I explored its use in what could be described as an audit process. If further developed, CoMFideS could help researchers and policy makers establishing service benchmarks for quality of care. As a self-monitoring tool, CoMFideS might be used to identify strengths or development needs, as well as helping in decision making processes regarding allocation of resources. Future studies might benefit from taking our findings forward to assess whether the measure could be used beyond the NHS.

Defining the context

One of the main learning outcomes I gained from conducting this research was that, to study fidelity, one must never lose sight of context. As such, it is important to reflect on the research process as a whole, and how the project took shape. Before even outlining the project, Prof Pilling and I had a series of thought-provoking conversations about my long term interest in global mental health research. While exploring some options we talked about how beneficial it would be for me to learn about psychotherapy research, service development, and the field of implementation research. While looking at some options, the ODDESSI trial came to mind. I had never heard of Open Dialogue nor had I received much training in systemic psychotherapies; however, something about the trial caught my attention.

The idea of implementing in the NHS a model which, in its origins, had been at such odds with the medical model of care seemed quite radical to me and, perhaps, it still does. I could not see how ideas of social justice and dialogism, or calls for fully individualized treatment designs without set contracts, plans, or pre-established endings could be sustained in the NHS. The model seemed intuitively sensible yet –I worried– perhaps overly optimistic. In a way, even from the outset of the project, I positioned myself as sceptical towards the model and this, perhaps, influenced my approach towards the research.

The systematic review was a useful process for understanding and contextualising this complexity. It helped me understand implementation evaluations and the role that fidelity measures have in defining and benchmarking complex interventions, such as the Open Dialogue approach. As Melissa Lotmore and I planned the scope of the review and noticed the overlap of concepts and definitions, I worried it could turn out to be too ambitious for a DClInPsy; however, if we were to develop reliable and effective measures of adherence and fidelity, it was crucial that we attempted to develop a framework that could be replicable in the future. Christopher Cooper proved to be an invaluable asset to this process. His knowledge on systematic search strategies allowed us to merge our search terms into an elaborate and replicable syntax that was sensitive to both our terms and could therefore capture potential overlaps. Even though our review revealed a wide variety of measures we would not have been able to identify otherwise, it still missed others that could have enriched our finding; nevertheless, the resulting review represents a step forward towards achieving a finer typology of fidelity measurement, and offers a more systematic approach to their design. It now rests on future studies to test its utility.

Dialogues and power

As I read about Open Dialogue, its history, and its preliminary findings, I began to think that something about the model could indeed be transported to the British health system. When I attended my first Open Dialogue conference in early 2018, I was fortunate to meet a group of passionate and thoughtful Open Dialogue practitioners who shared their experiences of using (and living) the dialogic model. Some of them had lived experiences of severe mental health difficulties and had been treated using a dialogic approach. Peers and practitioners seemed to share the view that *something* about tolerating the uncertainty that accompanies a mental health crisis, building genuine and curious interactions, and allowing for a dialogue within and between people –rather than a debate of conflicting viewpoints– produced a “new understanding” of difficulties.

Although certainly inspiring, I could not tell Open Dialogue apart from the Batesonian ideas from which the model emerged (Seikkula & Olson, 2003). Perhaps I still cannot. I wondered whether a reliance on network meetings and collaborative care actions meant that teams themselves needed to be structured in a certain way that allowed for flexible arrangements and immediate response to crisis to actually work (e.g. number of staff, caseloads, service links, etc.). If that were the case, then perhaps these features could be ascertained and eventually fostered in the few existing Open Dialogue teams in the country.

The idea of studying the organizational structures around Open Dialogue, although not a new one, was received with ambivalence. On one hand, the potential for developing a measure that could help Open Dialogue teams identify their strengths and areas where they could improve their practice seemed attractive. In fact, Professors Douglas Ziedonis, Mary Olson, and Jakko Seikkula (2015) had already developed a

draft for the ‘10 Organizational Principles of Open Dialogue’ which had highlighted some domains that could be important for effective dialogic practice (e.g. network working, collaboration, immediate support). On the other hand, however, the notions of ‘effectiveness’, ‘assessment’, ‘fidelity’, and ‘structure’, seemed to resonate with the institutionalised past against which Open Dialogue had once rebelled. As a consequence, conversations about Open Dialogue fidelity were usually approached with hesitation and –perhaps symmetrically– the same scepticism I once had.

On hindsight, this academic stalemate became an important source of learning about ‘openness’, ‘dialogue’, ‘power’, and most importantly, the notion of ‘schismogenesis’. In *Steps to an ecology of mind*, Bateson (1972) used the term schismogenesis to refer to a self-sustaining –and potentially destructive– intergroup dynamic in which “the behaviour of one party elicits a particular reaction in the other, which in turns reinforces the initial reaction of the former”. Bateson went further to add that these forms of interactions could potentially damage the system beyond repair if not addressed in due time. I began wondering whether the fact that I was approaching the problem of Open Dialogue fidelity as ‘*The Researcher*’ from ‘*The ODESSI Trial*’ already implied a certain hierarchy which may have echoed the psychiatric model of care. Understandably, this might have raised some resistance in practitioners, which increased my attempts to get answers, and so forth. A new strategy had to be developed to go beyond this impasse and, effectively, *open* a dialogue.

The value of anthropology

Learning about Open Dialogue was like learning about a new culture with a language of its own. In fact, throughout the research, I found myself working particularly close to anthropologists interested in the cultures behind different dialogic

interactions (e.g. network meetings, peer support workers, service-users, caregivers, etc.). Although I have never been trained in ethnographic approaches I began seeing parallels with my own research. After all, beyond the development of a measure, what the project was ultimately aiming to answer was regarding what made Open Dialogue a unique social networks approach to crisis when compared to the traditional model of care. To effectively study Open Dialogue, I had to change my initial strategy of studying the model *as an outsider* to immersing myself (as far as possible) in the Open Dialogue culture to understand it *from within*. In other words, what started as a monologic interaction (see ML's thesis) aimed at 'assessing' Open Dialogue principles, gradually became an 'exploratory' endeavour to understanding a seemingly different way dealing with distress.

As previously discussed, the findings suggest that CoMFideS might turn out to be a valuable tool for understanding multicomponent interventions for severe mental health difficulties. However, based on my 'ethnographic experiences' I believe the research could have greatly benefitted from a large qualitative component. In other words, studying psychosocial interventions cannot disregard both the contextual factors that surround them and relational dynamics within them. As discussed on Chapter 1, in order to understand the fidelity features of a model of care, it is also necessary to understand how its own microculture and internal narratives promote or obstruct their attainment. The item on 'Service structure and culture' aimed to address this, but was unable to go beyond the degree of model cohesiveness among practitioners. However, as discussed with anthropology experts, an ethnography component addressing both models (i.e. Open Dialogue and 'care as usual') might have revealed additional features that could have added nuance to some of our scoring anchors. Perhaps our fidelity

interviews, which could potentially be transcribed and qualitatively analysed in future studies, could enhance our understanding about what these model ‘ethea’ actually entail.

Final reflections: The scientist practitioner stance

As a scientist, I believe that one of the features that made this research process so enriching was, ironically, the flexibility, openness, and tolerance of uncertainty it required on its every stage. The measure development process was lengthy and, at times, it seemed like it would not reach fruition. The early bureaucracy, the language barriers, and the politics involved made this project a challenging one; however, although tentative, the results were promising. Likewise, it is important to acknowledge that this research would not have been fruitful had it not been for the cross-disciplinary relationships built and fostered throughout the process. The exchanges between professionals and academics during conferences, meetings, and informal conversations gave our findings a new depth. Lastly, given the close links between fidelity, adherence, and clinical practice, I sometimes had to use my own clinical experience and insight to address power dynamics and to tackle ambiguity while developing the scale. Although, indeed, this meant that additional efforts had to be made to manage risks for bias, the use of myself as tool for research also helped overcome some practical hurdles sometimes inherent to the research process itself.

On a similar note, this research also informed my role as a practitioner. First, I had the opportunity to learn about Open Dialogue, systemic practice, and their benefits for this client population. Second, I gained a wealth of knowledge about crisis and continuing community care teams and what makes them such an irreplaceable resource for building successful and sustainable mental health systems. This research tested my ability to reflect on how organizational dynamics can have an impact on service delivery

and how sometimes, real or perceived resource disparities (e.g. training, supervision, staffing) –rather than practitioner competence or dedication– may lead teams to drift away from their goals and models of care.

No matter the model of care, with this research I consolidated my views about the significance of (1) network working and network engagement, (2) tolerance of uncertainty, (3) genuineness, and (4) the fostering of dialogue and curiosity in clients *and* practitioners, in attaining and sustaining positive outcomes. I hope to carry this learning forward in my career as a clinical psychologist and to continue finding ways of connecting and empowering people and engaging the networks around them.

References

- Bateson, G. (1972). *Steps to an ecology of mind*. Chicago, IL: University of Chicago.
- Gaffney, C. (2012). Complex interventions for children and young people: exploring service delivery frameworks and characterising interventions (Doctoral Dissertation, University College London). Retrieved from <http://discovery.ucl.ac.uk/1367068/1/C.GAFFNEY%20VOLUME%201.pdf>
- Monroe-DeVita, M., Teague, G., & Moser, L. (2011). The TMACT: a new tool for measuring fidelity to assertive community treatment. *Journal of the American Psychiatric Nurses Association*, 17(1), 17-29. DOI: 10.1177/1078390310394658
- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P., Cooper, J., Eaton, J., Herrman, H., Herzallah, M., Huang, Y., Jordans, M., Kleinman, A., Medina-Mora, M., Morgan, E., Niaz, U., Omigbodun, O., Prince, M., Rahman, A., Saraceno, B., Sarkar, B., De Silva, M., Singh, I., Stein, D., Sunkel, C., & Unützer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157),

1553-1598.

Seikkula, J., & Olson, M. (2003). The open dialogue approach to acute psychosis: Its poetics and micropolitics. *Family process*, 42(3), 403-418.

Ziedonis, D., Olson, M., & Seikkula, J. (2015). *10 Organizational Criteria for Open Dialogue*. The University of Massachusetts Medical School. Worcester, MA.

Appendix A

Systematic search strategy

Appendix A.

Search strategy: Programme fidelity measures of complex mental health interventions for severe mental health difficulties

Search Syntax	Search Narrative
<ol style="list-style-type: none"> 1. (Mental\$ and health\$).ti,ab,kw,ot. 2. (((("Mental health" or psychiatr\$ or "community mental health") adj2 (service\$ or institution\$ or team\$)) or (communit\$ adj3 (treatment\$ or therap\$)) or (collaborat\$ adj3 care) or (multi\$ adj3 interven\$) or (famil\$ adj3 (treatment\$ or therap\$))).ti,ab,kw,ot. 3. 1 or 2 	<p>The specific focus on fidelity reflects the research question of the study.</p> <p>The truncation (i.e. \$) of interven\$ aims to identify intervention, intervene, inteventive, etc.</p> <p>.ti,ab,kw,ot. indicates that the search will be conducted in the following fields: title, abstract, and author generated keyword.</p> <p>Lines 1 and 2 aim to focus on “multi-component mental health interventions”. Line 1 focuses on the field of mental health. Line 2 focuses on different ways of describing multi-component mental health interventions.</p> <p>Line 3 combines the free text of line 1 and line 2 so that both are searched for.</p>
<ol style="list-style-type: none"> 4. fidel\$.ti,ab,kw,ot. 5. (adher\$ adj3 (measur\$ or metric\$ or referenc\$ or standard\$ or scal\$ or instrument\$ or assess\$)).ti,ab,kw,ot. 6. *Psychometrics/ and (adher\$ or consist\$ or reliab\$ or integrity).ti,ab,kw,ot. 7. (psychometr\$ and (adher\$ or consist\$ or reliab\$ or integrity)).ti,ab,kw,ot. 8. 4 or 5 or 6 or 7 	<p>Lines 4-8 aim to focus on “measures”. Line 4 focuses on the concept of fidelity. Line 5 aims to identify adherence measures specifically. Line 6 uses a FOCUS operator on the MeSH term for psychometrics, to direct the search to alternative ways of describing adherence and fidelity. Line 7 has the same rationale but broadening up the search to identify alternative forms of psychometric terms.</p> <p>Line 8 bring together the concepts set out in lines 4-7 ensuring that all identified concepts are searched for.</p>
<ol style="list-style-type: none"> 9. exp mental disorders/ 10. (Mental\$ and (disorder\$ or ill\$)).ti,ab,kw,ot. 	<p>Lines 9-12 focus on severe mental health difficulties. Line 9 explodes the MeSH term for ‘mental disorders’. Line 10 uses</p>

<p>11. (bipolar or ((feed\$ or eat\$) adj2 disorder\$) or "ED" or "depress\$" or "MDD" or "psychotic depression" or "depressive psychosis" or "personality disorder\$" or "PD" or "EUPD" or schizophreni\$ or "schizophrenia spectrum" or "psychotic disorders" or psychosis or "substance-related disorders" or "substance abuse" or "substance misuse" or "drug addiction" or "suicide" or "self-harm" or "conduct disorder" or "severe mental illness" or "serious mental illness" or "SMI").ti,ab,kw,ot.</p> <p>12. 9 or 10 or 11</p>	<p>alternative search terms for mental illness. Line 11 aims to identify specific conditions. We used a broad definition of severe mental health difficulties.</p> <p>Line 12 brings together the concepts set out in line 9-11 ensuring that all identified concepts are searched for.</p>
<p>13. 3 and 8 and 12</p>	<p>Line 13 links together the search terms for 'multi-component interventions' with those of 'measures' and 'severe mental health difficulties' to narrow down the search.</p>

Appendix B

Data extraction template

Appendix B.

Data extraction template

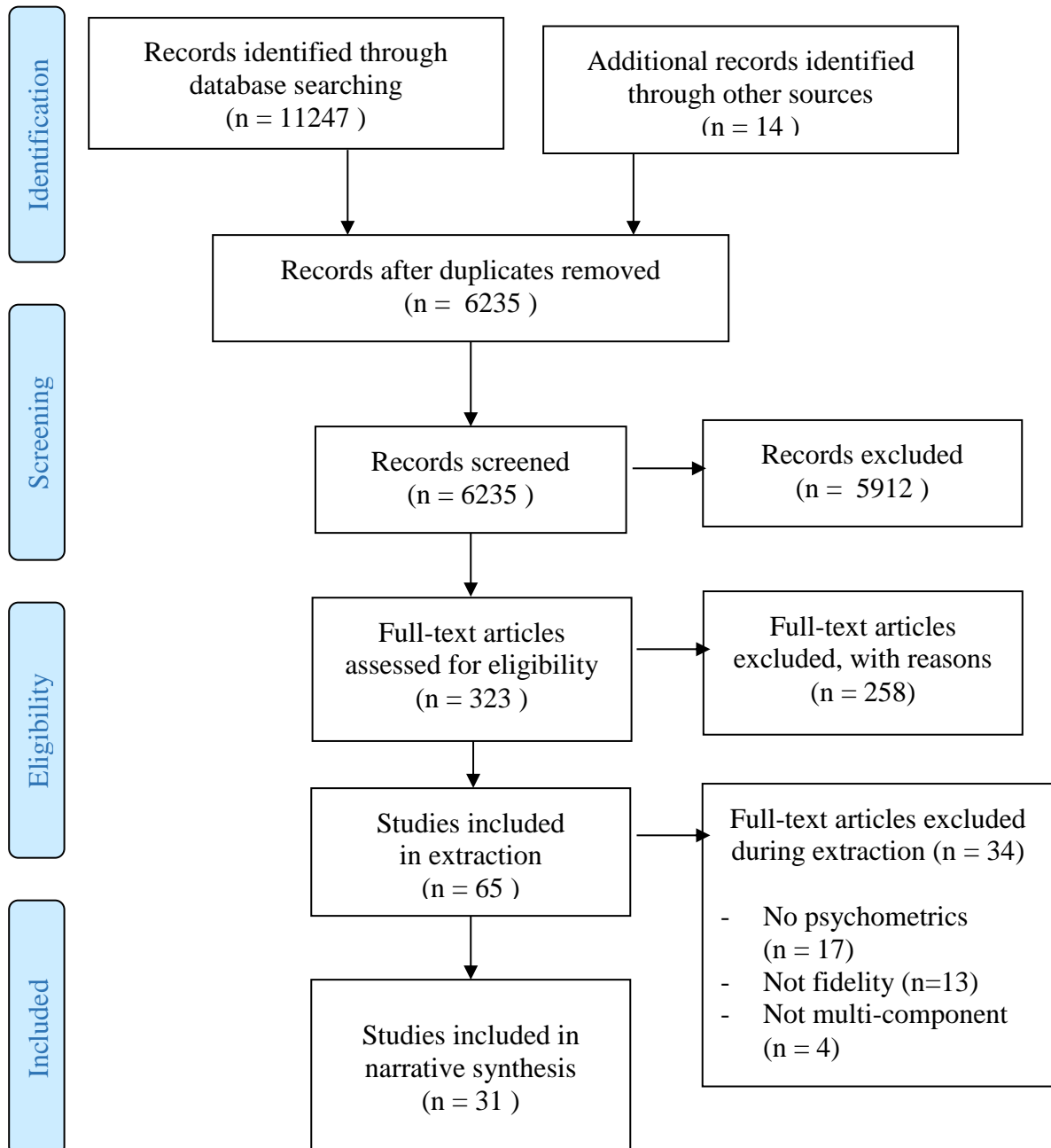
Reference	
Measure name	
Intervention	
Age Group	
Subscales	
Item count	
Rated by	
Scoring	
Cut-offs	
Participants	
Data sources	
Duration	
Reliability	
Validity	
Fidelity definition	
Training required?	
Manual available?	
Development method	
Notes	

Appendix C

PRISMA flow diagram

Appendix C.

PRISMA Flow Diagram.



Appendix D

Overview of programme fidelity measures

Appendix D. Overview of programme fidelity measures.

Measure	Item count	Subscales (Number of items)	Scoring and cut-offs
1. Adolescent Treatment Program (ATP) Environment Scale	21 items	1) Peers as gatekeepers (4) 2) Mutual help (5) 3) Enhancement of community belonging (5) 4) Contact with outside community (1) 5) Community/clinical management (6) 6) Level system (2)	Dichotomous scale (presence/absence) - Cut-offs not specified
2. Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services (CIMHRRS)	52 items	1) Program mission (4) 2) Program demographics & composition (10) 3) Organizational boundaries (5) 4) Program functioning (7) 5) Treatment team structure and process (10) 6) Assessment process (9) 7) Treatment planning (5) 8) Treatment provision (2)	1-5 Likert-type scale with behavioural anchors - Cut-offs not specified
3. Crisis Resolution Team Fidelity Scale (CORE CRT)	39 items	1) Referrals and access (10) 2) Content and delivery of care (16) 3) Staffing and team procedures (10) 4) Timing and location of care (3)	1-5 Likert-type scale with behavioural anchors - Cut-offs not specified
4. Dartmouth Assertive Community Treatment Fidelity Scale (DACTS)	28 items Originally 26: Staff size (HR11) and Role of consumers (S10) included in 2008 (McGrew et al., 2013)	1) Human Resources/Staffing patterns ⁸ (11) 2) Organizational boundaries/structure ⁸ (7) 3) Nature of services/service delivery ⁸ (10)	1-5 Likert-type scale with behavioural anchors - >4 full fidelity - 3-4 = moderate - < 3 = no fidelity

Appendix D (Continued). Overview of programme fidelity measures.

Measure	Item count	Subscales (Number of items)	Scoring and cut-offs
5. Dual-Disorder Treatment Fidelity Scale (DDT)	20 items	Unidimensional	1-5 Likert-type scale with behavioural anchors - Cut-offs not specified
6. Flexible Assertive Community Treatment Scale (FACTS)	60 items	1) Team structure (12) 2) Team process (12) 3) Diagnostics & interventions (13) 4) Organization of services (10) 5) Level of social services (5) 6) Use of routine outcome measures (3) 7) Level of professionalization (5)	1-5 Likert-type scale with behavioural anchors - Cut-offs not specified
7. Index of Fidelity of Assertive Community Treatment (IFACT)	17 items	1) Staffing (4) 2) Organization (7) 3) Service (6)	0-1 Minimum Thresholds Approach (1= criterion met or exceeded) - Cut-offs not specified
8a. Individual Placement and Support – 15 item scale (IPS-15)	15 items	1) Staffing (3) 2) Organization (3) 3) Service (9)	1-5 Likert-type scale with behavioural anchors - >65 = Consistent (\bar{X} =4.33) - 56-65 = Partial/Fair (\bar{X} =3.66) - <56 = Not IPS
8b. Individual Placement and Support Fidelity Scale (IPS)	15 items	1) Staffing (3) 2) Organization (3) 3) Service (9)	1-5 Likert-type scale with behavioural anchors - Cut-offs not specified

Appendix D (Continued). Overview of programme fidelity measures.

Measure	Item count	Subscales (Number of items)	Scoring and cut-offs
9. Individual Placement and Support – 25 item scale (IPS-25)	25 items	Not specified	1-5 Likert-type scale with behavioural anchors - Cut-offs not specified
10. Integrated Dual Disorders Treatment Fidelity Tool (IDDT)	26 items	1) Organization (12) 2) Treatment (14)	1-5 Likert-type scale with behavioural anchors - >4 = High - 3-4 = Moderate - < 3 = Low
11. Quality of Supported Employment Implementation Scale (QSEIS)	33 items	1) Vocational Staffing (6) 2) Organization (11) 3) Services (14 + 2 IPS items)	1-5 Likert-type scale - < 4.3 = Accurate - 4.0-4.3 = Moderate - 3.7-4.0 = Borderline - <3.7 = Low
12. Wraparound Fidelity Index (WFI)	44 items on the care facilitator and caregiver forms 32 items on the youth form	1) Family Voice and choice (4) 2) Youth and Family Team (4) 3) Community-based services (4) 4) Cultural competence (4) 5) Individualized, strength-based services (4) 6) Natural supports (4) 7) Continuation of care (4) 8) Collaboration (4) 9) Flexible resources (4) 10) Outcome-based services (4)	0-2 Likert-type scale - >85 = High - 80-84 = Above average - 75-79 = Average - 70-74 = Below average - 0-69 = Non-wraparound

Note: IPS = Individual Placement and Support Fidelity Scale

Appendix E

Programme fidelity assessment methods

Appendix E. Programme fidelity assessment methods.

Measure	Rated by	Data Sources	Duration	Training
1. Adolescent Treatment Program (ATP) Environment Scale	Service users	Activity logs (Self-report)	Once per week	No training available
2. Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services (CIMHRRS)	2 trained raters	<ul style="list-style-type: none"> • Semi-structured interviews <ul style="list-style-type: none"> ○ Employment specialists ○ Programme director • Policy and procedure manuals • Chart reviews • Internal agency documents 	16 hours	No manual available Training available Manual available
3. Crisis Resolution Team Fidelity Scale (CORE CRT)	3 trained raters	<ul style="list-style-type: none"> • On-site interviews <ul style="list-style-type: none"> ○ CRT manager ○ CRT staff team ○ Managers of associated services ○ 6 service users ○ 6 carers • 10 anonymised case records (latest consecutively discharged) • Service policies and records • Routinely collected data 	1 day	Training available No manual available

NOTE: CRT = Crisis and Recovery Team.

Appendix E (Continued). Programme fidelity assessment methods.

Measure	Rated by	Data Sources	Duration	Training
4. Dartmouth Assertive Community Treatment Fidelity Scale (DACTS)	2 trained raters	<ul style="list-style-type: none"> • Documentation <ul style="list-style-type: none"> ○ Programme authority ○ Responsibility ○ Policies ○ Procedures • Management IT systems <ul style="list-style-type: none"> ○ Staffing ○ Clientele ○ Services ○ Contacts • Interviews <ul style="list-style-type: none"> ○ Team leader(s) ○ Practitioners ○ Clients • Observation <ul style="list-style-type: none"> ○ Team meetings ○ Intervention (home visits) • Randomly selected clinical records • 9 survey tables (McGrew et al., 2013) 	1 day	Training available
	Self-report (McGrew et al., 2013)			Manual available
5. Dual-Disorder Treatment Fidelity Scale (DDT)	2 trained raters	<ul style="list-style-type: none"> • On-site interviews <ul style="list-style-type: none"> ○ Programme directors ○ Clinicians • Chart reviews 	Not specified	Training available (40 hours) No manual available

Appendix E (Continued). Programme fidelity assessment methods.

Measure	Rated by	Data Sources	Duration	Training
6. Flexible Assertive Community Treatment Scale (FACTS)	2 trained raters	Not specified	30-90 minutes (\bar{X} = 45 min)	Training available No manual available
7. Index of Fidelity of Assertive Community Treatment (IFACT)	Not specified	Pre-existing datasets	Not specified	Not specified
8a. Individual Placement and Support - 15 Item (IPS-15)	2 trained raters	<ul style="list-style-type: none"> • Telephone or on-site interviews <ul style="list-style-type: none"> ○ Program leaders ○ Employment specialists ○ Clients ○ Family members 	60 minutes	Training available (Systematic description of each item with examples) Manual available
8b. Individual Placement and Support Fidelity Scale (IPS)	2 trained raters Self-report (Becker et al., 2001)	<ul style="list-style-type: none"> • Interviews (including telephone; McGrew et al., 2005) <ul style="list-style-type: none"> ○ 1 programme leader ○ 2+ employment specialists ○ Clients • Observation <ul style="list-style-type: none"> ○ Meetings ○ Intervention • Client charts • Survey (Becker et al., 2001) 	1 day (90-120 minutes)	Training available No manual available

Appendix E (Continued). Programme fidelity assessment methods.

Measure	Rated by	Data Sources	Duration	Training
9. Individual Placement and Support - 25 Item (IPS-25)	2 trained raters	<ul style="list-style-type: none"> • Site visit • Interviews • Observation <ul style="list-style-type: none"> ○ Meetings ○ Community contacts • Client charts 	1.5 days	Training available Manual available
10. Integrated Dual Disorders Treatment Fidelity Tool (IDDT)	2-3 trained reviewers	<ul style="list-style-type: none"> • Interviews <ul style="list-style-type: none"> ○ Team leader ○ Practitioners ○ Clients • Observation <ul style="list-style-type: none"> ○ Meetings ○ Intervention ○ Client charts 	1 day	Training available Manual available
11. Quality of Supported Employment Implementation Scale (QSEIS)	2 trained raters	<ul style="list-style-type: none"> • Telephone interview <ul style="list-style-type: none"> ○ Programme directors 	90-120 min	Training available No manual available
12. Wraparound Fidelity Index (WFI)	1 or 2 trained raters	<ul style="list-style-type: none"> • Interviews (On-site or telephone) • Observation 	15-40 min per site	Training available Manual available

NOTE: CF = Care facilitator form; CG = Caregiver form; Y = Youth form

Appendix F

Programme fidelity measures (Psychometric properties)

Appendix F. Programme fidelity measures (Psychometric properties).

Measure	Reliability	Validity
1. Adolescent Treatment Program (ATP) Environment Scale	<i>Internal consistency (Faw, 2003):</i> Cronbach's alpha = .74 <i>Test-retest reliability (one week):</i> $r = .52 (p < .03)$ (Faw, 2003)	<i>Convergent validity:</i> Positively correlated with the Working Alliance Inventory (Horvath & Greenberg, 1994) $r = .44 (p < .04)$ (Faw, 2003)
2. Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services (CIMHRRS)	<i>Inter-rater reliability (Johnson, 2011):</i> ICC = .99 <i>Internal consistency (Cronbach's alpha) (Johnson, 2011):</i> Qualitative variables = .98 Program mission = .85 Organizational boundaries = .85 Team structure and process = .82 Assessment Process = .96 Treatment planning = .92	<i>Face and content validity:</i> Good. Extensive consultation with experts and panel guided refinement. <i>Construct validity:</i> Distinguishes between comprehensive PSR, intensive case management, maximum security SLP, residential PSR, and PSR day programme (Johnson, 2011).
3. Crisis Resolution Team Fidelity Scale (CORE CRT)	<i>Inter-rater reliability (Lloyd-Evans et al., 2016):</i> ICC = .97 (Range = .95-.98) <i>Internal consistency (Lloyd-Evans et al., 2016):</i> Cronbach's alpha = .97	<i>Face and content validity:</i> Excellent. Extensive consultation with experts (modified Delphi method), group-based statement conceptualization, concept mapping, cluster structure, interviewee feedback.
4. Dartmouth Assertive Community Treatment Fidelity Scale (DACTS)	<i>Inter-rater reliability:</i> IRR = .78 (Kidd et al., 2010) IRR = .44 (Winter & Calsyn, 2000) IRR = .59 (Johnsen et al., 1999) Human resources = .89 Organizational boundaries = .8 Nature of services = .58 ICC = .99 (McHugo et al., 2007) <i>(Continues in next page)</i>	<i>Face and content validity:</i> Very good. Informed by literature describing the model, expert consensus and previous research on ACT. <i>Construct validity:</i> Differentiates ACT from brokered case management. <i>Criterion/Predictive validity:</i> Predicts outcome of DACTS telephone version (Power=.81)

NOTE: ICC = Intraclass coefficient; PSR = Psychosocial rehabilitation programme; SLP = Social learning programme; ACT = Assertive Community Treatment.

Appendix F (Continued) Programme fidelity measures (Psychometric properties).

Measure	Reliability	Validity
4. Dartmouth Assertive Community Treatment Fidelity Scale (DACTS) (Cont.)	<p><i>Inter-rater reliability (Cont.):</i> ICC_{Phone} = .98 (McGrew et al., 2013) Human resources = .97 Organizational boundaries = .77 Nature of services = .97 ICC_{Self report} = .77 (McGrew et al., 2013) Human Resources = .47 Organization = .61 Nature of services = .86</p> <p><i>Internal consistency (Cronbach's alpha):</i> Total (26 items) = .92 (Teague et al., 1998) Human resources = .77 Organizational boundaries = .79 Nature of services = .83</p> <p><i>Test-retest reliability (Winter & Calsyn, 2000):</i> Stability = .28 No significant change over time (F(2, 11) = .80, $p > .05$, ES = .11)</p>	(See previous page)
5. Dual-Disorder Treatment Fidelity Scale (DDT)	<p><i>Inter-rater reliability (Wilson et al., 2009):</i> 92% agreement in chart reviews ICC = 0.95 ICC (across programmes) = .95 ICC (within programmes) = .77-.99 Item level ICC = .57-1.0</p>	<p><i>Construct validity:</i> Mixed results. The item level ratings demonstrated validity, but the analyses of total scores and implementation labels did not support the validity of conclusions drawn from these levels of data (Wilson et al., 2009).</p>
6. Flexible Assertive Community Treatment Scale (FACTS)	<p><i>Inter-rater reliability:</i> ICC = .88-.99 (Nugter et al., 2016)</p>	<p><i>Face and content validity:</i> Good (Based on DACTS and expert opinion)</p>

NOTE: IRR = Inter-rater reliability; ICC = Intraclass coefficient, ES = Effect size

Appendix F (Continued) Programme fidelity measures (Psychometric properties).

Measure	Reliability	Validity
7. Index of Fidelity of Assertive Community Treatment (IFACT)	<i>Internal consistency (Cronbach's alpha):</i> Total = .81 (McGrew et al., 1994) Staff = .50 Organization = .62 Service = .67	<i>Criterion/Predictive validity (McGrew et al., 1994):</i> Predicted a reduction in hospitalization days ($d_{Total} = 0.60$; $d_{Organization} = 0.56$; $d_{Staff} = 0.54$).
8a. Individual Placement and Support - 15 Item (IPS-15)	<i>Inter-rater reliability:</i> IRR = .80 (Bond et al., 1997) ICC = .20 (Bond et al., 2001) <i>Internal consistency (Cronbach's alpha):</i> Total = .92 (Bond et al., 1997) Staffing = .92 Organization = .65 Service = .29 Total = .80 (Bond et al., 2001)	<i>Face and content validity:</i> Modelled after IFACT and PACT. Items derived from IPS manual and brainstorming sessions among authors, drawing on experience on implementing this model and SE literature. <i>Construct validity:</i> <ul style="list-style-type: none"> • Kappa = 0.66 ($p < 0.001$) testing correspondence for fidelity labels (i.e. IPS = Consistent, Other SE = Partial consistency, Other VR = Not IPS) (Bond et al., 1997) • Distinguished IPS from SE ($d = 1.32$) and other VR ($d = 2.59$) (Bond et al., 2001)
8b. Individual Placement and Support Fidelity Scale (IPS)	<i>Inter-rater reliability:</i> ICC = .92 (.67-.99) (Bond et al., 2011) <i>Internal consistency:</i> Cronbach's alpha = .27 (.30-.49) (McGrew et al., 2005)	<i>Face and content validity:</i> Format and assessment procedure follows the conventions from DACTS. <i>Criterion/Predictive validity:</i> <ul style="list-style-type: none"> • Scores correlates with mean competitive employment ($r = .76$, $p < .01$; Becker et al., 2001).
9. Individual Placement and Support - 25 Item (IPS-25)	<i>Internal consistency:</i> Cronbach's alpha = .88 (Bond et al., 2012)	<i>Criterion/Predictive validity:</i> <ul style="list-style-type: none"> • Scores correlated with employment rates ($r = .34$; Bond et al., 2012) Services that sustained fidelity had better employment rates (Bond et al., 2016)

NOTE: IRR = Inter-rater reliability; ICC = Intraclass coefficient; IPS = Individual Placement and Support; SE = Supported employment; VR = Vocational rehabilitation; IFACT = Index of fidelity of assertive community treatment; PACT = Program in assertive community treatment model scale.

Appendix F (Continued) Programme fidelity measures (Psychometric properties).

Measure	Reliability	Validity
10. Integrated Dual Disorders Treatment Fidelity Tool (IDDT)	<i>Inter-rater reliability:</i> $F(3,34) = 1.07$ ($p = .38$) (Harrison et al., 2017) ICC = .89 (McHugo et al., 2007)	<i>Face and content validity:</i> Good. Informed by literature describing the model and extensive consultation with experts and panel guided refinement.
11. Quality of Supported Employment Implementation Scale (QSEIS)	<i>Inter-rater reliability:</i> IRR = .8 (.63-.96) (Bond et al., 2002) ICC = .93-.97 (McGrew et al., 2005) <i>Internal consistency (Cronbach's alpha):</i> $\alpha = .72$ (Bond et al., 2000) $\alpha = .51$ (Bond et al., 2002) $\alpha = .54$ (.12-.80) (McGrew et al., 2005)	<i>Construct validity:</i> <ul style="list-style-type: none"> • Correlated with IPS-15 scores ($r = .85$, $p < .05$) (Bond et al., 2000) • Correlated with IPS scores ($r = .97$, $p < .001$) and with Supportive Employment Consultation and Training (SECT) adherence rating ($r = .47$, $p < .05$) (McGrew et al., 2005) <i>Criterion/Predictive validity:</i> Predicted employment rates ($r = .42$, $p < .05$) (McGrew et al., 2005)
12. Wraparound Fidelity Index (WFI)	<i>Inter-rater reliability (ICC):</i> Total = .51 (Pullman et al., 2013) Total = .58 (Bruns et al., 2008). CF-CG = .44 CG-Y = .49 CF-Y = .45	<i>Face and content validity:</i> Excellent. Use of literature, Delphi method for expert consensus, and three revisions of the measure to date <i>Criterion/Predictive validity:</i> <ul style="list-style-type: none"> • Score inversely correlated with youth needs ($r = .44$, $p < .05$; $\beta = -.14$, $p < .05$) (Effland et al., 2011). • Baseline scores correlated with child behavioural strengths six months later ($r = .79$, $p < .05$) (Effland et al., 2011).

NOTE: IRR = Inter-rater reliability; ICC = Intraclass coefficient; IPS = Individual Placement and Support Fidelity Scale; IPS = Individual Placement and Support -15 item scale

Appendix F (Continued) Programme fidelity measures (Psychometric properties).

Measure	Reliability	Validity
12. Wraparound Fidelity Index (WFI) (Cont.)	<p><i>Internal consistency (Cronbach's alpha):</i></p> <p>Care facilitator = .78 (Bruns et al., 2004); Care facilitator = .82 (Bruns et al., 2008) Caregiver = .90 (Bruns et al., 2004) Caregiver = .91 (Bruns et al., 2008) Youth = .88 (Bruns et al., 2004) Youth = .84 (Bruns et al., 2008) CF₂₀₀₇ = .67, CF₂₀₀₈ = .76 (Kernan, 2014) CG₂₀₀₇ = .84 CG₂₀₀₈ = .90, CG₂₀₀₉ = .89 (Kernan, 2014) Y₂₀₀₇ = .83, Y₂₀₀₈ = .90, Y₂₀₀₉ = .84 (Kernan, 2014)</p> <p><i>Test-retest reliability (2-week; Bruns et al., 2008):</i></p> <p>$r_{CF} = .84$ $r_{CG} = .88$ $r_Y = .64$</p>	(See previous page)

NOTE: IRR = Inter-rater reliability; ICC = Intraclass coefficient; CF = Care facilitator form; CG = Caregiver form; Y = Youth form

Appendix G

Definitions of fidelity

Appendix G. Definitions of fidelity.

Measure	Definition of fidelity
1. Adolescent Treatment Program (ATP) Environment Scale	Extent to which treatment programs adhere to the principles of a therapeutic community model (Faw, 2003).
2. Comprehensive Inventory of Mental Health Recovery and Rehabilitation Services (CIMHRRS)	Characterization of process differences in service settings for people with serious mental health difficulties considering the structural and organizational components in addition to functional processes of assessment and treatment provision (Johnson, 2011).
3. Crisis Resolution Team Fidelity Scale (CORE CRT)	Measure of implementation of intervention or programme models (Lloyd-Evans et al., 2016).
4. Dartmouth Assertive Community Treatment Fidelity Scale (DACTS)	<ul style="list-style-type: none">- Conformity with prescribed elements and the absence of non-prescribed elements (Johnsen et al., 1999).- Adherence to the principles and procedures specified in the evidence-based practice models (McHugo et al., 2007).- Measure of adherence to a model but also a tool to provide specific feedback to reinforce strengths and improve areas of weakness (Rollins et al., 2010).- Careful specification of the critical components of a model based on operational definitions, which prevents drifting away from the model and allows to evaluate the respective contributions of theoretically distinct components (Teague et al., 1998).- Measuring the degree to which the treatment program was implemented as designed (Winter & Calsyn, 2000).
5. Dual-Disorder Treatment Fidelity Scale (DDT)	Robustness of program implementation consisting of system, program, and clinician level assessments (Wilson et al., 2009).
6. Flexible Assertive Community Treatment Scale (FACTS)	Level of implementation of the programme model (Nugter et al., 2016).
7. Index of Fidelity of Assertive Community Treatment (IFACT)	Conformity with prescribed elements and absence of non-prescribed elements of a programme model (McGrew et al., 1994).

Appendix G (Continued). Definitions of fidelity.

Measure	Definition of fidelity
8a. Individual Placement and Support - 15 Item (IPS-15)	<ul style="list-style-type: none"> - Adherence to standards of a programme model, used to empirically identify its critical ingredients (Bond et al., 1997). - Degree to which a specific program meets the standards for a program model (Bond et al., 2001).
8b. Individual Placement and Support Fidelity Scale (IPS)	<ul style="list-style-type: none"> - Adherence to the principles and procedures specified in the evidence-based practice models (McHugo et al., 2007). - Accurate representation of critical program components (Becker et al., 2001). - Adherence to evidence-based programme models or the extent to which an intervention or practice is implemented as intended at the system, organization, program, practitioner, or client level (Bond et al., 2011). - Measure of implementation of a programme model (McGrew et al., 2005).
9. Individual Placement and Support - 25 Item (IPS-25)	Adherence to the evidence-based principles of a programme model (Bond et al., 2012, 2016).
10. Integrated Dual Disorders Treatment Fidelity Tool (IDDT)	Adherence to the principles and procedures specified in the evidence-based practice models (McHugo et al., 2007).
11. Quality of Supported Employment Implementation Scale (QSEIS)	Degree of implementation of a practice or degree of attainment of practice standards (Bond et al., 2000).
12. Wraparound Fidelity Index (WFI)	<ul style="list-style-type: none"> - Adherence to the core principles of a therapeutic model (Bruns et al., 2004, 2008). - Quality of a therapeutic model as delivered to service users or adherence to the elements of a therapeutic model (Bruns et al., 2006). - Extent to which communities apply the principles of a therapeutic model in services (Effland et al., 2011) - Adherence to the principles and primary activities of a therapeutic model or the extent to which both the principles and activities of an intervention are implemented in service delivery (Kernan, 2014). - Measurement of a team's adherence to a well-defined therapeutic process (Pullmann et al., 2013).

Appendix H

HRA Ethics Approval

Professor Stephen Pilling
Director of the Centre for Outcomes Research and
Effectiveness
UCL
CORE, 1-19 Torrington Place
WC1E 7HB

Email: hra.approval@nhs.net

26 February 2018

Dear Professor Pilling

Letter of HRA Approval

Study title:	Open Dialogue: Development and Evaluation of a Social Network Intervention for Severe Mental Illness (ODDESSI)
IRAS project ID:	233483
REC reference:	18/LO/0026
Sponsor	NELFT

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from the [HRA website](#).

Appendices

The HRA Approval letter contains the following appendices:

- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval

The document “*After Ethical Review – guidance for sponsors and investigators*”, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:

- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the *After Ethical Review* document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the [HRA website](#), and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the [HRA website](#).

Scope

HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found through [IRAS](#).

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application

IRAS project ID	233483
-----------------	--------

procedure. If you wish to make your views known please use the feedback form available on the [HRA website](#).

HRA Training

We are pleased to welcome researchers and research management staff at our training days – see details on the [HRA website](#).

Your IRAS project ID is **233483**. Please quote this on all correspondence.

Yours sincerely

Thomas Fairman
HRA Assessor

Email: hra.approval@nhs.net

Copy to: *Ms Natasha Clarke, UCL, (Sponsor Contact)*
Dr Sandeep Toot, North East London NHS Foundation Trust, (Lead NHS R&D Contact)

Appendix I

The Community Mental Health Fidelity Scale (CoMFideS)

Community Mental Health Fidelity Scale (CoMFideS)

Alvarez-Monjaras, M. & Pilling, S.

This **Community Mental Health Fidelity Scale (CoMFideS)** is designed to measure the programme fidelity of the Open Dialogue and standard NHS crisis and community services care. The scale addresses four aspects of service provision:

1. Service structure and culture
2. Access
3. Delivery of intervention
4. Community linkage and support.

Additionally, an Open Dialogue addendum evaluates the extent to which Open Dialogue services as a whole support and enact specific dialogical operational principles.

<input type="checkbox"/> CAU
<input type="checkbox"/> OD

<input type="checkbox"/> Managers
<input type="checkbox"/> Practitioners

TEAM: _____ **TRUST:** _____

DATE: ____/____/____ **RATED BY (INITIALS):** _____

SERVICE INFORMATION

1. Development and operation (please see 5)

- Stand-alone
- Integrated with other services

2. Staffing

- Number of temp staff in last 24 months: _____
- FTE staff members: _____
- Psychiatrists
- Support Workers
- Nurses
- Peer-support Workers
- Nurse Assistants
- Advocates
- Psychologists/Psychotherapists
- Other (specify): _____
- Occupational Therapists
- _____
- Social Workers

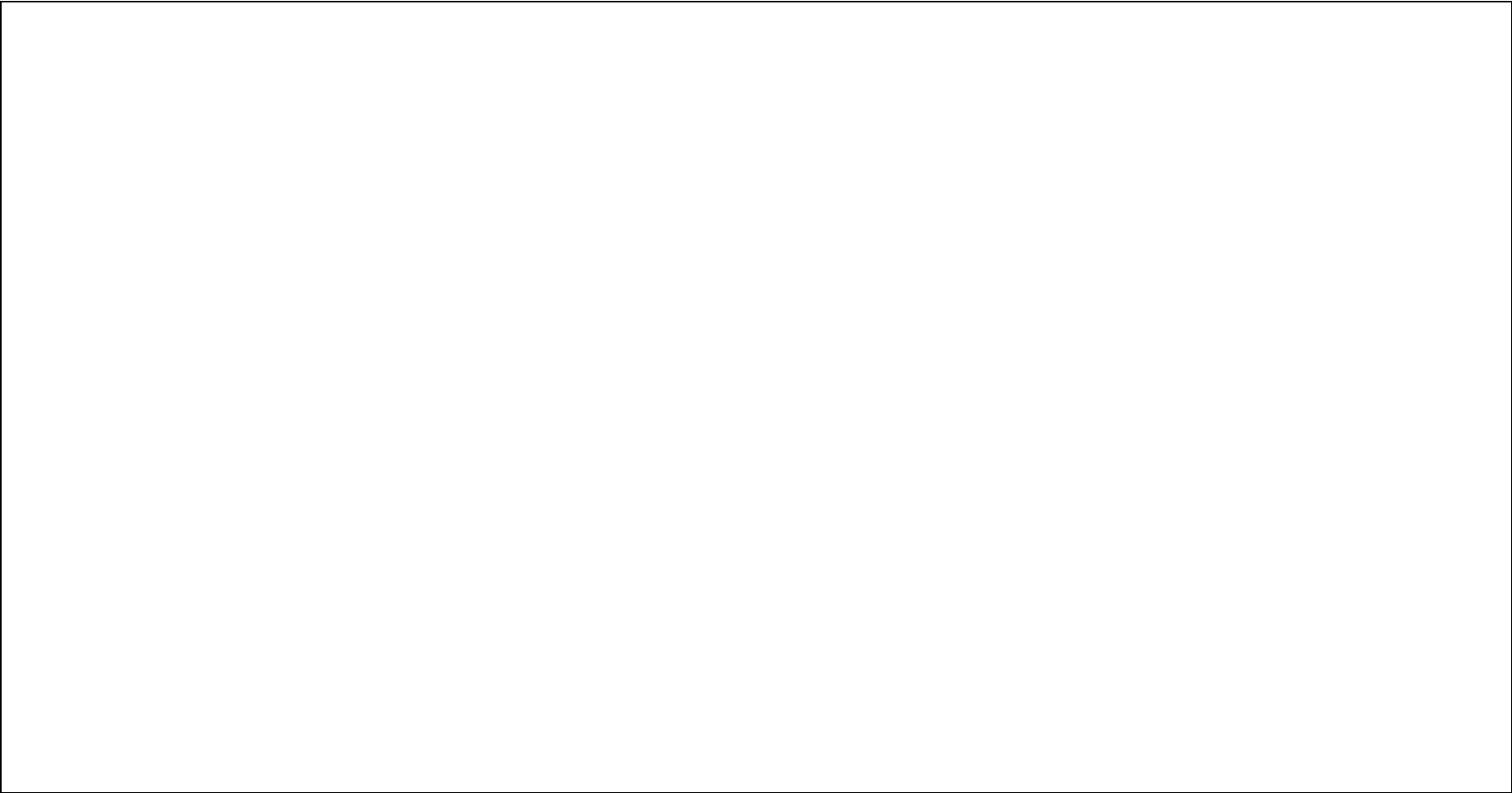
3. Team Caseload:

- Current: _____
- Maximum for service: _____
- Maximum for individual: _____
- Individual average: _____

4. Team and Individual Supervision:

- Team meeting frequency (day/week): _____
 - Assessment and referral
 - Operations/Management
 - CPA
- Supervision frequency and duration:
 - Individual: _____
 - Group: _____
 - Other (specify): _____
 - _____
 - _____

5. Organizational structure map (please include points of access and relevant pathways in and out of the service)



SERVICE STRUCTURE AND CULTURE

Dimension	1 Not clearly present	2 Somewhat present	3 Mostly Present	4 Fully present
1. Service ethos and comprehensiveness - Comprehensive, well-articulated and shared view of the model of care provided. - Service has a person-centred approach to care.				
2. Staff training - Staff are trained and competent for their role. - Training needs are monitored and needs responded to				
3. Supervision - Operational policy includes a clear supervision model. - All team members receive regular supervision.				
4. Staff roles - Composition of the service is consistent with core service functions. - Operational policy includes clearly defined roles and responsibilities.				
5. Service capacity - Service staffing is appropriate to cover the case load. - Knowledge and skills available to effectively deliver key service requirements.				
6. Routine outcome monitoring - Service actively monitors service-user outcomes and has a culture of responsiveness to feedback. - Operational policies include clear guidance and benchmarks to assess and improve service outcomes.				
7. Safety - Operational policy has clear and explicit risk management and safeguarding policies. - Evidence of concern for the safety of service-users and staff members, openly discussed within the team.				
8. Service-user involvement in co-production - There is service-user involvement in the development and planning of the service. - There is service-user involvement evaluation, and assessment of the service.				

ACCESS AND ENGAGEMENT

Dimension	1 Not clearly present	2 Somewhat present	3 Mostly Present	4 Fully present
9. Access to the service - Referral criteria (i.e. population served and sources) are explicit - Clear and structured care programmes/pathways (e.g. assessment, content of care, discharge)				
10. Providing information - Service provides clear information about treatment plans and models of working. - Service provides clear information about relationship to other relevant services				
11. Prompt action - Protocols are in place with expected response times including for urgent and emergency cases - Service has the capacity to respond within the policy time limits.				
12. Identification of support systems - Service-users' support systems/networks are identified and engaged with where possible. . - Service-users' support systems/networks are included in care planning and provision where possible.				
13. Flexibility of response - Service provides or supports access to an appropriate range of interventions to address service-users' needs - Service provides or supports access to appropriate social, educational, housing or employment services				
14. Assertive engagement - Service makes all possible efforts to engage the service-user even when initial contact is problematic - Service makes all possible efforts to engage significant others, professionals and institutions				

DELIVERY OF CARE

Dimension	1 Not clearly present	2 Somewhat present	3 Mostly Present	4 Fully present
15. Continuity of care - Staff involved in the initial meetings have a key, coordinating role throughout an episode of care - When service-users are re-referred , services endeavour to allocate staff previously involved in their care				
16. Establishing clinical meetings - Meetings are convened by a staff member and an agenda is set collaboratively with the service-user. - Meetings take place in appropriate settings (e.g. home visits, community settings) where possible of the service-user's choosing				
17. Collaborative decision making - Decisions regarding care and treatment are developed in collaboration with the service-user. - Where decisions are made without the service-user (e.g. to ensure safety) team members are aware of such decisions inclusive decision making process.				
18. Information-sharing and communication - Patient records and letters are either summarised or written collaboratively - All correspondence will be copied to the service-user, other than in exceptional circumstances.				
19. Service-user involvement in delivery of care -Evidence of service-user involvement in supporting the planning of care (e.g. advocacy, etc.) - Evidence of service-user involvement in supporting the provision of care (e.g. volunteering, peer support.)				
20. Coordination of care - There are effective systems in place to ensure the proper coordination of care (e.g. advocacy, etc.) - Coordination of care is monitored and reviewed through appropriate service structures				

COMMUNITY SUPPORT AND ENGAGEMENT

Dimension	1 Not clearly present	2 Somewhat present	3 Mostly Present	4 Fully present
21. Service linkage - Service promotes effective collaboration between mental health and other care services - Collaboration and coordination between mental health and other care services is reviewed and monitored				
22. Community links (Practitioner level) - Care, follow-up and liaison support to access community services are effectively provided - service-users are supported in accessing community services				
23. Community links (Support system) - Assessment of the service-user's support system/network capacity to support community engagement - Active engagement with the service-user's support system/network to enhance and strengthen links				
24. Carer involvement and support - Family/carer's needs and support are considered and addressed. - The service provides information about local support services for families and carers (e.g. carers groups, welfare advice, child support).				
25. Discharge and aftercare - Care coordination/ or network meetings include discussion and agreement of end of care, including referrals to other health and social care services. - Discharge meetings/discussions involve the service-user's social support system or network.				

OPEN DIALOGUE ADDENDUM

Dimension	1 Not clearly present	2 Somewhat present	3 Mostly Present	4 Fully present
1. Transparency - All discussions about the service-user and their network occur with them present. - There is a culture of ‘nothing about them, without them’ and neither the service-user nor members of their network are talked about when they are not present.				
2. Self-disclosure - Professionals share their own lived experiences (self-disclosure) in both intervision and network meetings if deemed relevant and appropriate.				
3. Intervision frequency - All members of the team meet at least weekly for intervision (or group supervision)				
4. Intervision content and structure - Intervision focuses on adherence to the key principles of Open Dialogue and the clinicians’ own emotions and reactions, while minimising content of the actual case wherever possible. - Intervision includes team members sharing personal reflections in pairs/groups, reflections on which are then shared with team members.				
5. Team self-work - A regular programme of self-work is on-going within the team where such work is engaged in by all team members - Team members are encouraged to maintain a regular individual self-work practice				
6. Open Dialogue training - All team members have completed or are undergoing an 0accredited Open Dialogue training.				
7. Open Dialogue continuing professional development (CPD) - Team members attend regular (at least annually) CPD delivered by accredited Open Dialogue trainers.				

Appendix J

CoMFideS Manual

Community Mental Health Fidelity Scale (CoMFideS)

Rating Manual

by Mauricio Alvarez-Monjaras & Prof. Stephen Pilling

June 2019

Do not copy, distribute, or reproduce without the authors' consent. This file is part of the Open Dialogue: Development and Evaluation of a Social Network Intervention for Severe Mental Illness (ODDESSI) programme (NIHR grant: RP-PG-0615-20021).

Introduction

The **Community Mental health Fidelity Scale (CoMFideS)** is designed to describe the structure, functioning, pathways, community links and delivery of care provided by crisis and recovery community mental health services. It is designed as a measure of programme fidelity of both standard NHS crisis care and best dialogic practice. To this end CoMFideS addresses four key areas:

1. **Service structure and culture (8 items)** – This section is concerned with the overall service ethos and comprehensiveness of the model of care (including model consistency and coherence), staff training, supervision, staff roles, service capacity, routine outcome monitoring, safety practices (for service-user and staff), and the degree to which service users get involved in service co-production (i.e. development and evaluation of services).
2. **Access and engagement (6 items)** – This sections looks into referral and treatment pathways, provision of information about the service (to referring agencies and service users), prompt action, the proactive and effective inclusion of the service-user's support systems, flexibility of response (i.e. range of interventions available), and assertive engagement strategies.
3. **Delivery of care (6 items)** – This section covers continuity of care (i.e. the extent to which the same care coordinator is maintained throughout a treatment pathway), the way in which clinical appointments or network meetings are arranged and convened, collaborative decision making, information-sharing and communication practices, coordination of care, and the degree to which service users get involved in the provision of care.
4. **Community linkage and support (5 items)** – This section is concerned with the extent to which the service is embedded in the immediate community (i.e. the amount and strength of links to other services and support agencies) and practitioners' skilfulness in effectively using these resources (e.g. liaison, follow-up, inclusion to meetings), and engaging the service-user's own support systems (instead of deploying the services' own resources). This section also focuses on the extent to which caregivers are taken into consideration and supported throughout treatment and the nature and sustainability of discharge and aftercare practices.

Additionally, the **Open Dialogue addendum (7 items)** evaluates the extent to which

Open Dialogue teams support and enact specific dialogic principles into practice. This section assesses how the principle of transparency is enacted in the team, the attitude of the team towards self-disclosure, the team's frequency and content of intervention (i.e. the model's take on group supervision), team self-work practices, the promotion of certified Open Dialogue training courses, and the presence of a coherent Open Dialogue continued professional development programme.

Administration

CoMFideS was designed to be administered and rated by two people. This is because community mental health services are quite complex in nature and thus having two raters would make the data collection and rating process more reliable, valid, and efficient, than if done by a single person. Likewise, it would allow for rating inconsistencies to be discussed by the raters and, in so doing, ensure a more realistic and balanced score. Therefore, it is expected that at least one of the raters is experienced in working with community mental health services.

Considering the complexity of community mental health services, data collection must come from a variety of sources. This includes a review of policies and procedures, and interviews with senior staff and front line practitioners.

Policies and procedures may include any form of written material that describes the service in as much detail as possible, such as the operational policy (including policies on clinical risk management and safeguarding, supervision, and training), referral and service protocols, audit reports, and annual reports. This written material will allow the raters to get a preliminary understanding of the service as a whole and might reduce interview times, as it will allow the raters to fill in as much information about the service beforehand and identify areas where more information might be needed.

The data collection process and scoring involves a series of steps, as summarized below:

I. Initial contact with service

The first point of contact will be by mail. A letter explaining the scale and its purpose will be sent to the service, requesting the following:

- a. Copies of all relevant operational policies and procedures

- b. Information on organizational arrangements, frequency and content of meetings, full time equivalent (FTE) and support staff numbers (including permanent, temporal, and leaving staff).

Additionally, the letter will include a request to arrange interviews with with: (a) two team leaders or senior staff, and (b) two front line practitioners.

II. Documentation review

In order to make the rating procedure a more efficient one, prior to the assessors visiting the service to meet the staff, raters will review the documentation provided. This information will help populate the measure and cover as many areas as possible, thus shortening the time needed for face-to-face interviews. Ratings during this stage will be provisional and can be modified based on further information obtained during interviews or where additional documentation is provided. A key purpose of this review stage is to identify gaps in information and to guide the questions to staff and team leaders.

III. Fidelity interviews

The core function of the interviews is to ensure that all 32 items of the scale (25 in the case of non-Open Dialogue teams) are taken into consideration from both ends of the staffing spectrum. Therefore, although interviews will be somewhat similar, the precise nature and structure of the questions will be primarily influenced by the information provided by the service. In some cases, the questions will only seek to confirm or clarify aspects of the measure; in some other cases –where not enough information is available– further information may be sought. Ideally, all interviews would be expected to take place during the same day and take no more than 60 minutes each.

IV. Final service rating

As the interviews are undertaken and any missing information is collected, both raters will simultaneously and independently rate a copy of the CoMFideS. Once finished, they will review and reach a consensus in the ratings. Where a consensus is not possible, assessors may need to seek further information or clarification from the service, or seek advice from a senior colleague.

Scoring

CoMFideS is rated on a scale of 1 (Not clearly present) to 4 (Clearly present), yielding to a final overall score of 100. In the case of Open Dialogue teams, fidelity to the model will be assessed on a 1 to 28 points. On all cases a score of one indicates that the principle at hand is not present or there is insufficient evidence of its enactment in the team's way of functioning, whereas a score of 4 indicates that the principle is enacted or carried out in an excellent manner and with no visible shortcomings or inconsistencies across the team.

Service structure and culture

1. Service ethos and comprehensiveness

Services with more consistent and coherent working ethos are better prepared to address problems and complex situations (Walker, 2003). Therefore, high quality services are expected to have (1) a comprehensive, well-articulated, and shared view of the model of care provided. This would not only be expected on a policy level but also at the staff level, where (2) all team members should be aware of the working ethos, expectations, and procedures, and (3) be engaged in promoting and sustaining model of care. Likewise, (4) high quality services are expected to have a priority given to service-user-defined goals and values (i.e. person-centredness) rather than to service resources or to goals identified by the professionals (i.e. resource-orientation).

4 points

All four premises are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

There are some minor disagreements in terms of the services' ethos and model of work (e.g. confusion over certain goals or the approach).

2 points

Team members are aware of the service vision and ethos, but there are some clear splits in terms of its enactment (e.g. clear disagreements or disputes about the model of care).

1 point

The service lacks a clear ethos and has no clear or shared values or model of care.

2. Staff training

Best services are expected to show evidence of appropriate training (Patel et al., 2018). In that sense, high quality services would be those where (1) all team members are trained and competent for their roles, including (2) both basic and specialized training (depending on each job description). Likewise, (3) training requirements and expectations should be clearly described in relevant policies and (4) high quality services would be expected to show evidence of providing in-service training (including induction training) for all team members.

4 points

All four elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

There is evidence that all team members are appropriately trained, with some disagreements between interviewees.

2 points

Staff is appropriately trained for their roles but with a lack of in-service training and limited monitoring systems put in place to ensure the staff remains competent for their roles.

1 point

Staff is undertrained or not competent for their role (e.g. inadequate training background or a mismatch between training and job description), and there are no provisions for in-service training.

3. Supervision

Consistent, congruent, and goal-oriented supervision is at the heart of any good clinical practice (Roth & Pilling, 2008); as such, high quality services are expected to show evidence of (1) a comprehensive approach to supervision in their operational policy. (2) All service members should be receiving regular supervision (i.e. at least

once per month). Furthermore, (3) it would be expected that the provided supervision is congruent with the therapeutic models being used.

4 points

All three elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Most staff members attend regular supervision (i.e. once per month) but sometimes relying on ad-hoc supervision.

2 points

Supervision arrangements are not comprehensive (e.g. fortnightly, superficial, or not congruent with the therapeutic model being used) and staff relies primarily on ad-hoc arrangements.

1 point

Supervision is intermittent/ad-hoc or takes place less frequently than bi-monthly.

4. Staff roles

Effective services need appropriate staffing levels (Gilbody, Bower, Fletcher, Richards, & Sutton, 2006); therefore, to score highly in this item, a service should be able to demonstrate that (1) its composition is consistent with core service functions (i.e. a good alignment between service core-functions and job roles and responsibilities), and (2) professional, theoretical, and technical diversity among the staff (e.g. psychiatrists, psychologists, nurses, occupational therapists, social workers, support workers, peer-support workers, etc.). Furthermore, high quality services would be expected not only to (3) have well-defined roles and responsibilities, but also evidence that (4) these roles are understood by all team members.

4 points

All four elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

There is an alignment between service core-functions and job roles/responsibilities.

Job roles/responsibilities are also well-defined; however, there is uncertainty in the interviewee responses over role boundaries.

2 points

Certain service functions, job roles and/or responsibilities (e.g. overrepresentation of social workers over nurses) are somewhat prioritized and there is a clear confusion over role boundaries or considerable overlap between roles.

1 point

There is a clear priority given to certain job roles and service functions, a concerning shortage of permanent positions, and the staff is unclear about their functions and role boundaries (leading to considerable clinical time waste and inefficiency).

5. Service capacity

Services should be capable of providing the best possible care for all service users (Patel et al., 2018); therefore, it is paramount that staffing levels are appropriate to cover the case load. This item refers primarily to staffing numbers and caseloads; therefore, highest scores would be given to services where (1) the clinician-to-service-user ratio is appropriate to the tasks, frequency of contacts, and frequency of reviews, and (2) waiting times are reasonable, and (3) there is a balance between permanent and temporal positions.

4 points

Both interviewees agree there is an appropriate clinician-service-user ratio and evidence of professional, theoretical, and technical diversity among the staff.

3 points

The clinician-service-user ratio is found to be somewhat unmanageable but interviewees agree that they have procedures in place to maintain reasonable waiting times.

2 points

The clinician-service-user ratio is found to be unmanageable and interviewees agree that more staff is needed to manage the service caseload.

1 point

The clinician-service-user ratio is clearly unmanageable, waiting lists are unacceptable and there is a clear imbalance between permanent and temporal positions.

6. Routine Outcome Measurement

Service development and best practice are both dependent on consistent, reliable, and meaningful service monitoring practices (British Psychological Society, 2018; Patel et al., 2018). Therefore, it would be expected that high quality services would be those that (1) explicitly seek feedback about the quality of service practices, procedures, and outcomes (e.g., outcome measures and satisfaction data), in order to identify areas of opportunity for development and potential service oversights. Likewise, services should be capable of providing evidence that their operational policies include (2) clear guidance and benchmarks to assess and improve service quality and, what is more, they would also be expected to (3) take action to address quality concerns by implementing meaningful changes without shifting from the overarching service ethos.

4 points

All three elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

There is evidence of routine outcome monitoring practices (e.g. symptom trackers, goals based outcomes, etc.) and interviewees are clear about both service evaluation guidelines and benchmarks; however, there is little evidence of feedback-driven change (i.e. a process, data collection efforts, and outcomes).

2 points

Routine outcome monitoring practices (e.g. symptom trackers, goals based outcomes, etc.) are not consistent and interviewees are unclear about service evaluation guidelines and benchmarks for feedback-seeking practices.

1 point

There is little or no evidence of routine outcome monitoring practices, service evaluation or feedback-seeking practices (including efforts to address the feedback).

7. Safety

Since the publication of *The Five Year Forward View for Mental Health* (Mental Health Taskforce, 2016) there has been a clear push for NHS England to make mental health services safer (Tingle, 2019). Therefore, it is of utmost importance that (1) services are clear about their safety protocols and procedures. (2) This is not only in relation to service-user safety (e.g. safeguarding, care plans, communication, risk assessments, care coordination), but also for team members themselves (e.g. lone-worker policies, availability of alarms, safety calls, joint community visits, etc.).

4 points

There is enough evidence and descriptions of both features of safety, with minor inconsistencies or disagreement between interviewees.

3 points

Interviewees are clear about safety protocols and procedures (even with some minor disagreements) and there is evidence of well-established and enacted service-user safety

procedures; however, interviewees may consider their safety is not properly taken into consideration (e.g. not enough alarms, not enough safety calls).

2 points

Interviewees are clear about safety protocols and procedures and there is evidence of appropriate service-user safety procedures (albeit with clear disagreements between interviewees); however, interviewees may provide accounts of feeling unsafe.

1 point

Interviewees are unclear about safety protocols and procedures and consider that both service users' and their safety is not properly taken into consideration.

8. Service-user involvement in co-production

According to the Care Act 2014 statutory guidance (Department of Health, 2014), "Co-production" takes places when an individual influences the support and services received, (...) and the way that services are designed, commissioned and delivered". As such, services are expected to show evidence of service-user involvement in the development, planning, evaluation, and/or assessment of the service as a whole.

4 points

There is clear evidence of service-user involvement in the development, planning, evaluation, and/or assessment of the service as a whole (all four elements are present).

3 points

There is evidence of service-user involvement in the development, planning, evaluation, and/or assessment of the service as a whole (at least three of these elements are present).

2 points

There is evidence of service-user involvement in the development, planning, evaluation, and/or assessment of the service as a whole (at least two of these elements are present).

1 point

There is no evidence of service-user involvement in the development, planning, evaluation, and/or assessment of the service as a whole (one or none of these elements are present).

Access and engagement

9. Access to the service

To function effectively, services need (1) clear referral criteria (such as population served, thresholds and sources of referral), and that (2) clear and structured treatment pathways. Furthermore, (3) these pathways would ideally span from the point of acceptance, to describing the progress and transition through the service, to the point of discharge and follow-up.

4 points

All three elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Referral criteria and treatment pathways are clear, well-structured, and explicit in the operational policy, and there is clear information about treatment plans and models of working (including roles and contact details of practitioners).

2 points

Treatment pathways and referral criteria are not clear or explicit in the operational policy, there is unclear information about treatment plans and models of working, and there is insufficient contact information for service-users to seek assistance.

1 point

Referral criteria and treatment pathways are unclear or not present in the operational policy, unclear information about treatment plans and models of working, insufficient contact information for service-users to seek assistance, and there are no service adherence protocols put into place.

10. Providing information

Integrated care –including that of community mental health teams– is heavily reliant on clear communication between services, agencies, and service users, so as to avoid ‘inappropriate’ referrals and unmanageable workload pressures (Lester, Glasby, & Tylee, 2004). As such, it is important that services (1) make active efforts in educating local referrers (e.g. GPs, A&E, third-sector agencies) about their remit and inclusion/exclusion criteria. Likewise, clear information about in-house services and models of work available should be easily accessible to (2) external referrers and (3) service users.

4 points

All three elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Clear information about in-house services and models of work is easily accessible to external referrers and service users; however, there is little evidence of proactive education of local referrers about the service remit and referral criteria.

2 points

Information about in-house services and models of work may be available for external referrers and service users but this may not be evident or might be difficult to access. There is also little evidence of proactive education of local referrers about the service remit and referral criteria.

1 point

Information about in-house services and models of work is unavailable and no efforts are made to educate local referrers about the service remit and referral criteria.

11. Identification of support system

Evidence suggests that the most effective community mental health teams are those with more collaborative models of working (Patel et al., 2018). As such, services are expected to (1) arrange case coordination or network meetings for all service-users contacting the service. It is best if (2) all individuals, professionals and institutions in each service-user's social support systems or networks (e.g. family, friends, colleagues, social workers, carers, school teacher, counsellors, etc.) directly involved in the service-user's current problem are proactively identified. Doing this would provide the service with the capacity to (3) engage effectively with service-users and their support systems/networks (e.g. to arrange for transportation, tele-conferences, and/or home visits).

4 points

Both interviewees agree that the service makes proactive efforts (i.e. from the point of access) to identify a service-user's support network and openly invites those deemed relevant by the service-user.

3 points

There are some disagreements over the service's level of proactivity for identifying a service-user's support network; however, the network model of care is evident in the interviewees descriptions.

2 points

Even if identified, the service-user's support systems are rarely invited or included or they are only identified when deemed relevant for the service-user.

1 point

Service-user's support systems are seldom identified and very rarely includes for the treatment.

12. Prompt action

Duration of untreated psychosis (DUP) is an important predictor of negative short and long term outcomes (e.g., poor general symptomatic outcomes, more severe

positive and negative symptoms, lower likelihood of remission, poor social functioning, worse global outcomes) (Penttilä, Jääskeläinen, Hirvonen, Isohanni, & Miettunen, 2014). As such, high quality services are expected to provide evidence of (1) expected response time frames in their operational policy, as well as of (2) clear protocols for dealing with urgent and emergency cases. Further, it would be expected that services (3) actually have the capacity to respond within the policy time limits.

4 points

All three premises are clearly observed (i.e. no inconsistencies or disagreement between interviewees), response times are ideally within 24 hours, and face-to-face meetings take place within a week.

3 points

Urgent and emergency protocols are clear and the services established contact (at least by telephone) during the first 24 hours; however, face-to-face meetings can take more than a week to take place.

2 points

Interviewees are unclear about expected response time frames and the service is unable to respond (even by telephone) within 24 hours; however, face-to-face meetings tend to take place within a two-week period.

1 point

The operational policy has no accounts of expected response time frames or urgent and emergency protocols, the service takes more than 48 hours to respond (even by telephone), and face-to-face meetings can take more than two weeks to take place.

13. Flexibility of response

Services should be able to provide a wide range of interventions and services in order to adapt to the service-user's needs (Alanen, Lehtinen, Rökköläinen, & Aalatonen, 1991; Patel et al., 2018). High quality services would therefore be expected to (1) be able to provide or support access to as many interventions as needed to attend the service-user's needs (i.e. pharmacological, social, psychological, psychoeducation, peer-support interventions). Likewise, best services should be able to demonstrate (2) effective collaboration and coordination between the crisis and recovery elements of community care.

4 points

All two elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

There is a wide variety of services and interventions available and, when an intervention is not available, the team is efficient in establishing the required links.

2 points

The range of interventions is based on service resources rather than service-user need or there is a clear priority given to crisis management rather than recovery care.

1 point

There is an unclear evidence of a range of interventions available, lack of consideration for service-user needs (e.g. a set care package with the same interventions for all service-users), and a priority given to crisis management rather than recovery care.

14. Assertive engagement

Some service-users may find it difficult to engage with services or even approach them for a range of reasons (Kaufman, McDonell Cristofalo, & Ries, 2012).

Therefore, assertive engagement refers to the service's attempts to respond to these reasons and should make all possible efforts to include and engage significant others, professionals, and institutions already involved in the service-user's care and support, when initial contact is problematic. This may include behavioural, motivational or liaison strategies for which evidence should be provided.

4 points

Evidence of assertive engagement is clearly described, including multiple engagement strategies (e.g. home visits, contacting GP, liaising with other services).

3 points

There is evidence of assertive engagement using multiple engagement practices but interviewees have some disagreements about their level of proactivity in doing so.

2 points

There is some evidence of assertive engagement, however, engagement strategies

might be limited and purely related to liaison (e.g. just sending a second appointment letter and calling the GP).

1 point

There is no evidence of assertive engagement with service-user that are difficult to engage (e.g. no follow-up after a few calls, discharging them when no contact has been made, or relying solely on a DNA policy).

Delivery of care

15. Continuity of care

Continuity of care has been found to be important for mental health services (Patel et al., 2018; Bergström et al., 2018). Continuity of care not only promotes stronger treatment alliances but has also been suggested to promote faster recovery from crisis (Green et al., 2008, 2013). Therefore, high quality services would be expected to provide evidence that (1) all service-users have a named staff member coordinating his or her care throughout their treatment and (2) whenever a transfer of care is necessary, arrangements are made to maintain at least one of the original practitioners involved in the service-user's care. More importantly, this principle also applies to when a service user is re-referred to the service, in which occasion (2) the service should make active efforts to re-assign the original key worker involved in the service user's care

4 points

There is clear evidence of ensuring continuity of care during treatment and re-referral.

3 points

There is clear evidence of ensuring continuity of care during treatment, with most service user's having a named worker and a consistent group of specialists proactively managing their care throughout the whole treatment sequence; however, the service struggles to arrange for continuity of care in the context of re-referral.

2 points

There is some evidence of continuity of care during treatment, however, arrangements for continuity during re-referral rely primarily on service resources.

1 point

There is no consistency in care providers or staff involved in network meetings during treatment or re-referral.

16. Establishing clinical meetings

Clinical (or network) meetings are the backbone of community mental health service provision (Patel et al., 2018). Therefore, meetings should (1) be consistent with the model of service delivery (as stated on the operational policy). Likewise, best services should be trying their best to embed their helping efforts as near as possible to everyday life, circumstances, and relationships of the service-users; therefore – where possible– (2) these meetings should take place wherever the service-users finds it most convenient (e.g. home, community, hospital, etc.).

4 points

Both elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Clinical meetings are usually established based on service-user's wishes and take place wherever the service-user finds it most convenient; however, when unable to accommodate, the service tends to take the lead on the final decision about the location of the meeting.

2 points

Clinical meetings take into consideration the service-user's wishes but take place mostly inside the service's premises.

1 point

Clinical meetings do not take into consideration the service-user's wishes usually take place within the service premises, determined by practitioners.

17. Collaborative decision making

Shared decision making has become a growing area of interest, due to its theorized impact on clinical outcomes (Shay & Lafafa, 2015). Literature suggests that

promoting a horizontal (i.e. dialogical) rather than a vertical (i.e. directive) hierarchy increases trust in service providers and promotes agency, both of which are important factors for recovery (Laugharne, Priebe, McCabe, Garland, & Clifford, 2012). Therefore, services should provide evidence of (1) working with treatment plans that adapt to individual service-user's or their network's needs and requests. Likewise, services should ideally (2) be capable of refraining from making decisions on behalf of the service-user or their networks; (3) the service seeks to involve all relevant people in the decision-making process (e.g. service-user, family, GP, other services) and must therefore provide evidence of such practices (e.g. meeting minutes, action plans, care notes, etc.). Further, (4) a positive risk-taking attitude is to be encouraged, always making all possible efforts to ensure adequate safety for the service-user. Finally, (5) shared decision-making policies should be explicit in the operational policy.

4 points

All five elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Practitioners work within flexible treatment plans that adapt to individual service-user's or their network's needs, they refraining from making decisions on behalf of the service-user or their networks (involving all relevant people in the decision-making process), and a positive risk-taking attitude is encouraged; however, interviewees might have some disagreements as to the degree of flexibility of the team.

2 points

Practitioners work within somewhat flexible treatment plans that can adapt to individual service-user's or their network's needs; however, practitioners might be described as more "directive" (sometimes forgetting to involving all relevant people in the decision-making process). A positive risk-taking attitude is encouraged but practitioners tend to take the lead when discussing risk planning

1 point

The service has inflexible treatment plans that disregard the service-user's changing

needs and requests, service providers are directive or prescriptive in the decision-making process, and there is no evidence of a positive risk-taking attitude (even if the service-user's safety is being taken into consideration).

18. Information sharing and communication

Clear, transparent, and efficient communication is essential for any effective service (Borrill, West, Shapiro, & Rees, 2000). Therefore, high quality services are expected to make efforts towards this goal; for instance, services should (1) summarize and routinely spread the main themes of each meeting and (2) inform whether any decisions have been made. Likewise, whether or not a decision about care has been made, (3) efforts should be made to ensure that all people involved are aware and understand the next steps. Furthermore, if a person is not present in a meeting, (4) steps should be taken to communicate the nature and outcome of the meeting to them and relay their comments to future meetings. This not only helps for future reference, but also for quality assurance and supervision purposes. Finally, in order to make sure information-sharing practices are consistent and clear, (5) the relevant pathways and mechanisms should be included and described in the operational policy.

4 points

All five elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Information is summarized, communicated, understood, and relayed to all those involved in care in most cases but information-sharing protocols and mechanisms are not fully understood by interviewees.

2 points

The information is summarized, communicated, understood, and relayed to all those involved in care in some cases, but (a) information is not relayed to unavailable members or (b) most interactions between services take place behind-the-scenes, without the service-user's awareness (e.g. liaison with other teams), thus leading to problems in decision making and communication.

1 point

Information is neither summarized, nor communicated, understood, or relayed to all

those involved in care, leading to a centralized management of information and to accounts of confusion and problems in decision-making processes.

19. Service-user involvement in delivery of care

Recent literature suggests there are benefits to involving peers –or people with personal experiences of mental health difficulties– in using their lived experiences to support, advocate or provide care for other service users during their recovery process (Gillard & Holly, 2014). In line with this, services should be able to provide evidence of including service-users in the provision of care. This may be in the form of (in-house) volunteering, peer support, advocacy, teaching, leading groups, etcetera.

4 points

There is clear evidence of service-user involvement in the provision of care, in the form of a formal and structured arm of service provision (e.g. two or more peer support workers).

3 points

There is some evidence of multiple instances of service user involvement in the provision of care but without a clear model of work.

2 points

There is little evidence of service-user involvement in the provision of care (e.g. only one peer support worker) or, when present, it is mostly via referral to external panels and agencies.

1 points

There is no evidence of service-user involvement in the provision of care.

20. Coordination of care

Effective collaborative care is intimately dependent on effective care coordination (i.e. structure management plans, proactive follow-up practices, enhanced inter-professional communication) (Wagner, Austin, Von Korff 1996; Ramanuj, & Pincus, 2019). Services are thus expected to ensure (1) that all treatment plans and transfers of care to other services are proactively managed, with (2) reliable and efficient systems put into place to monitor treatment plans and agreed actions (e.g. shared records, IT systems, etc.). Finally, when possible (3) best services would be expected

to include external providers (e.g. external psychotherapists, social workers, employment agencies) to be involved in their shared service user's network or care coordination meetings.

4 points

All three elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Treatment plans and transfers of care to other services are proactively managed and external providers are included into network/care coordination meetings; however, these procedures might not always be entered or monitored thorough service records/IT systems.

2 points

Treatment plans and transfers of care to other services are not always proactively managed nor entered on monitoring systems; however, there are some attempts at involving external providers in network/care coordination meetings.

1 point

Treatment plans and transfers of care to other services are not proactively managed nor entered on monitoring systems. External providers are not included in network/care coordination meetings.

Community linkage and support

21. Service linkage

Well-connected and embedded services should be at the forefront of community mental health (Patel et al., 2018). Best services would therefore be expected to (1) be well embedded in their catchment areas, with multiple (and strong) links with various community resources and agencies. (2) Efforts to contact external or support agencies to inform them about the service's ethos would also be expected.

4 points

Both elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

External and support agencies are well informed about the service ethos (with enough evidence of different diffusion strategies) and staff takes action to help the service-user liaise with the wider network; however, liaison protocols are not explicit in the operational policy.

2 points

Even if the operational policy is explicit about liaison protocols and staff takes action to help the service-user liaise with the wider network, there is not enough evidence that external and support agencies are well informed about the service ethos.

1 point

Operational policy is not explicit about liaison protocols, external and support agencies are not informed about the service ethos and service provision only takes place in the service's premises.

22. Community links (Practitioner level)

All throughout the care pathway, practitioners are expected to (1) provide service users with follow-up on their progress and (2) liaison support to access health social care services. In that sense, (3) evidence of efforts to include external agencies in care coordination and network meetings would be appropriate. Likewise, practitioners would be expected to (4) provide effective support with basic living needs, benefits and debts, and/or urgent legal and social problems.

4 points

All four elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Practitioners provide follow-up on progress and provide liaison support to access community services. There is also evidence that the practitioners providing enough support with basic living needs, benefits and debts, and/or urgent legal and social problems; however, agencies are not always included in the care coordination process or into network meetings.

2 points

Practitioners provide follow-up on progress and provides liaison support to access community services for some cases. However, support with basic living needs, benefits and debts, and/or urgent legal and social problems are usually outsourced by signposting service-users to external agencies.

1 point

Practitioners do not provide enough follow up or liaison support and there is not enough evidence of efforts to include external agencies or of supporting with basic living needs, benefits and debts, and/or urgent legal and social problems.

23. Community links (Support system)

Enabling and empowering families and support systems to manage crises and difficulty is a key component for recovery promotion and maintenance of positive outcomes (Cochran, 1987; Jorm, 2012; Patel et al., 2018). This is why, beyond the local agencies and resources, practitioners ought to (1) displace the locus of control from the service to the network. In other words, practitioners should be able to (2) find means of engaging the service user's support systems and wider network (e.g. extended family, schools, friends, neighbours) to mutually support each other and promote the recovery process. For instance, instead of allocating service resources to accompany a service user to a GP visit, it would be preferable that someone from the service-user's network stepped forward to do so (without the need of coercion).

4 points

Practitioners are effective in engaging and enabling the service users' networks and support systems to carry out increasingly larger aspects of the treatment and recovery plan.

3 points

Practitioners are effective in engaging and enabling the service users' networks and support systems but interviewees may differ in terms of how centralized the responsibility for recovery is located (i.e. more in the service than in the network).

2 points

Practitioners struggle to engaging and enabling the service users' networks and

support systems or might have clearly differing views about the relevance of this feature to their model of care.

1 point

The service does not engage nor enabling the service users' networks and support systems.

24. Caregiver involvement and support

Caregiver wellbeing is essential for adequate support. Evidence suggests that caregiver burnout is associated with poorer outcomes and loss of care (Sharma, Sharma, & Pradhan, 2018; Whitlock, Lloyd-Richardson, Fisseha, & Bates, 2018). Therefore, it would be expected that (1) the service actively seeks to involve family and/or caregiver's in a service-user's care and support, and (2) their needs and support plan must always be considered and shared with service-user during sessions. Furthermore, it is essential that (3) services provide information about local support services (e.g. carers groups, welfare advice, child support), for which (4) a clear carer support policy and resources would be ideally provided.

4 points

All four elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

The service actively seeks to involve family and/or caregivers and considers their needs as part of a carer's support plan, as well as providing them with information about local support services, in most service-users; however, they don't provide carers with enough information about local support services.

2 points

The service considers a service-users' family and/or caregivers needs if and when requested by service-users (or only by performing a carer's assessment with no further action).

1 point

The service does not actively seek to include family and/or caregivers and there is little evidence of carer's support plans being developed.

25. Discharge and aftercare

Treatment in severe mental health difficulties does not usually end with full discharge. This is because the risk of relapse, network breakdown, and further complications can be expected (Drake & Whitley, 2014). Therefore, (1) clear discharge plans and aftercare policies and protocols would be expected, and best services would be those which (2) include discussion and agreement of end of care – including referrals to other health and social care services– in their care coordination or network. Discharge plans would (3) ideally involve the service-user’s social support system or network, rather than reliance on service structures, for which evidence (e.g. signed agreements, progress notes with alternatives explored, action plans, etc.) should be provided. Finally, (4) some efforts on following up on discharge plans would be expected.

4 points

All four elements are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

End of care is discussed and agreed upon during care-coordination or network meetings, including the service-user’s support system (rather than sole reliance on service structures) in the discharge plans; however, there is little follow-up once the service user has been discharged.

2 points

End of care is discussed and agreed upon during care-coordination or network meetings; however, these rely primarily on service structures rather than the service-user’s support system.

1 point

End of care is based on an agreed number of sessions or is seldom discussed or agreed upon with the service-user, no matter whether discharge plans rely on service structures or the service-user’s support system.

Open Dialogue addendum

26. Transparency

The cornerstone of Open Dialogue relates to the “openness” or transparency of the therapy planning and decision-making process (Olson, Seikkula & Ziedonis, 2014). As such, Open Dialogue services are expected to (1) have all discussions about the service-user’s treatment plan (e.g. hospitalization, medication, and treatment alternatives) while everyone is present in the room. Further, Open Dialogue services are expected to (2) have a consistent “nothing about them, without them” philosophy (Patel et al., 2018) that is shared and enacted by all staff members.

4 points

Both elements of transparency are clearly described, with minor inconsistencies or disagreement between interviewees.

3 points

Most discussions about service users occur with them present, however, interviewees may have some disagreements about the extent to which they enact this principle

2 points

Many discussions about service users take place when they ARE NOT present, for instance, with multiple conversations about them during team meetings or care planning discussions.

1 point

Talking about service users when they ARE NOT present is a common practice in the team.

27. Self-disclosure

Evidence behind the benefits of therapist self-disclosure in treatment outcomes is slowly growing (Danzer, 2018). Self-disclosure can also be an example of transparency in the room (Holmesland, Seikkula, & Hopfenbeck, 2014). As such, Open Dialogue services would be expected to have a clear positive and encouraging attitude towards clinical staff members sharing, where relevant, their own lived experiences in both intervision/supervision and with service-users during network meetings.

4 points

Staff members are expected to share their own lived experiences in both intervision/supervision and network meetings, where relevant.

3 points

Staff members are expected to share their own lived experiences in intervision/supervision OR network meetings BUT there is a tendency to either (a) oversharing or disclosing in inappropriate moments, or (b) some team members being reluctant to self-disclose.

2 points

Staff members discouraged from sharing their own lived experiences in intervision/supervision OR network meetings, but there is some degree of flexibility.

1 point

Self-disclosure is clearly discouraged in both intervision/supervision or network meetings.

28. Intervision frequency

Ideally, 20-30 minutes should be set off for Open Dialogue intervision (i.e. Open Dialogue's take on group supervision) ideally on a daily basis (Hopfenbeck, 2018). However, understanding this may not be possible for all services, it would be expected that intervision (or group supervision) takes place at least once per week with all team members joining.

4 points

Intervision (or group supervision) takes place at least weekly.

3 points

Intervision (or group supervision) takes place less than weekly but more than fortnightly.

2 points

Intervision (or group supervision) takes place fortnightly.

1 point

Intervision (or group supervision) takes place once a month or does not take place.

29. Intervision content and structure

Open Dialogue intervision tries to incorporate the values, intention and practice of the dialogical approach (Hopfenbeck, 2018). Therefore, additional to supporting clinical staff to work around therapeutic processes –including the clinician’s own emotions and reactions to clinical material– the main aim of intervision is to (1) help clinicians stay adherent to the Open Dialogue model, rather than the actual content of the sessions. Further, intervision sessions can be expected to have some level of flexibility; nevertheless, there are four core features that must be present in order to maintain fidelity to the Open Dialogue model (Hopfenbeck, 2018): (2) Every intervision starts with a brief mindfulness practice; (3) Team members share personal reflections in pairs/groups on the unique challenges they are experiencing in their interaction with the particular families/networks they are seeing; (4) Reflections are observed and then reflected on by the remaining team members (i.e. ‘fish bowl’ model); and (5) Original pair/group share a final reflection at the end.

4 points

Intervision includes all four key structural components (i.e. mindfulness, practitioner reflections, ‘fish bowl’, final reflection). Likewise, there is evidence of a team focus on adherence and practitioners’ own emotions and reactions, instead of content of the actual cases.

3 points

Intervision includes at least three key structural components (i.e. mindfulness, ‘fish bowl’, and reflections) and there is a focus on adherence and practitioners’ own emotions and reactions, however, while interviewees may consider difficulties in staying away from content from the actual cases the team is mindful of these deviations and tries to correct this.

2 points

Less than three key structural components of intervision are present. Although there is a focus on adherence and clinician’s own emotions and reactions, there is a tendency of discussing content from the actual cases and little team efforts to notice and correct this.

1 point

There is little focus on adherence and practitioners’ own emotions and reactions, and

a clear focus on content from the actual cases.

30. Team self-work

Therapeutic relationships in Open Dialogue are thought to be dependent on practitioners' ability to be fully present, open, and genuine, which are all qualities dependent on self-knowledge and self-development (Hopfenbeck, 2015). Therefore, additional to each practitioner's preferred self-work practices (e.g. mindfulness, psychotherapy, meditation, yoga, personal diary, etc.), Open Dialogue teams would be expected to (1) have an in-house regular and on-going self-work programme where family-of-origin, genogram, or other related work takes place. (2) Annual self-work retreats could also be encouraged, using training or study leave allowance where possible. Otherwise, (3) it can be beneficial for Open Dialogue teams to encourage all clinical staff to engage in and maintain their own personal self-work practice

4 points

There is a clear self-work programme and evidence of active encouragement for all team members, including training/study leave allowance.

3 points

There is a team self-work programme and evidence of active encouragement for all team members, but training/study leave arrangements are unclear or inflexible.

2 points

The service is open to external self-work teaching and/or workshops but there is no service-led programme and they are not included in training/study leave arrangements.

1 point

There is no scope for self-work teaching/training as part of the service's standard practice.

31. Open Dialogue training

In order to properly compare treatments in outcome studies, it is important that therapists share the same training, so as to reduce the amount of clinical variance. As such, high-fidelity Open Dialogue teams would be expected to be those where all

clinical staff has completed or is undergoing a recognised Open Dialogue training programme (see <http://open-dialogue.net/training/> for a list of recognised courses).

4 points

All clinical staff has completed or is undergoing a recognized Open Dialogue training programme.

3 points

All clinical staff –with a small number of exceptions (e.g. a couple of members of staff who have recently joined, but are expecting to start training soon)– has completed or is undergoing a recognized Open Dialogue training.

2 points

The majority of the clinical staff has completed or is undergoing a recognized Open Dialogue training, with most of the rest being due to be trained soon.

1 point

Less than half of the clinical staff has completed or is undergoing a recognized Open Dialogue training.

32. Open Dialogue continuing professional development (CPD)

Continuing professional development (CPD) is important for clinical staff to keep their skills and knowledge up to date, in order to practise safely and effectively (HCPC, 2017). Additional to the HCPC requirements, it is expected that Open Dialogue service staff attend at least an annual CPD day organised and delivered by recognised Open Dialogue trainers.

4 points

There is an annual CPD day organised and delivered by recognised Open Dialogue trainers AND all team members are actively encouraged to attend.

3 points

There is an annual CPD day organised and delivered by recognised Open Dialogue trainers but practitioners are not actively encouragement or supported to attend.

2 points

There is an annual CPD day that may not always be organised and/or delivered by recognised Open Dialogue trainers or there is no consistency in the model of training

provided.

1 point

There are no annual CPD days.

Key references

- Alanen, Y., Lehtinen, K., Rääköläinen, V., & Aaltonen, J. (1991). Needadapted treatment of new schizophrenic patients: experiences and results of the Turku Project. *Acta Psychiatrica Scandinavica*, 83(5), 363-372.
- Bergström, T., Seikkula, J., Alakare, B., Mäki, P., Köngäs-Saviaro, P., Taskila, J., Tolvanen, A. & Aaltonen, J. (2018). The family-oriented open dialogue approach in the treatment of first-episode psychosis: Nineteen-year outcomes. *Psychiatry Research*, 270, 168-175.
- Borrill, C., West, M., Shapiro, D., & Rees, A. (2000). Team working and effectiveness in health care. *British Journal of Healthcare Management*, 6(8), 364-371.
- British Psychological Society (2018). Code of Ethics and Conduct. London: BPS.
- Cochran, M. (1987). Empowering families: An alternative to the deficit model. *Social intervention: Potential and constraints*, 105-119.
- Danzer, G. (2018). Research of efficacy and outcomes. In G. Danzer (Ed.). *Therapist Self Disclosure: An evidence-base guide for practitioners*. London: Routledge.
- Department of Health. (2014). *Care and Support Statutory Guidance Issued under the Care Act 2014*.
- Drake, R., & Whitley, R. (2014). Recovery and severe mental illness: description and analysis. *The Canadian Journal of Psychiatry*, 59(5), 236-242.
- Gilbody, S., Bower, P., Fletcher, J., Richards, D., & Sutton, A. J. (2006). Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Archives of Internal Medicine*, 166(21), 2314-2321.
- Gillard, S., Holley, J., Gibson, S., Larsen, J., Lucock, M., Oborn, E., Rinaldi, M., & Stamou, E. (2015). Introducing new peer worker roles into mental health

- services in England: Comparative case study research across a range of organisational contexts. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(6), pp. 682-694.
- Green, C., Perrin, N., Leo, M., Janoff, S., Yarborough, B., & Paulson, R. (2013). Recovery from serious mental illness: Trajectories, characteristics, and the role of mental health care. *Psychiatric Services*, 64(12), 1203-1210.
- Green, C., Polen, M., Janoff, S., Castleton, D., Wisdom, J., Vuckovic, N., Perrin, N., Paulson, R., & Oken, S. (2008). Understanding how clinician-patient relationships and relational continuity of care affect recovery from serious mental illness: STARS study results. *Psychiatric Rehabilitation Journal*, 32, 9-22.
- Health and Care Professions Council (2017). *Continuing professional development and your registration*. London: HCPC.
- Hopfenbeck, M. (2015). Peer-supported Open Dialogue. *Context*, 138, 29-31.
- Hopfenbeck, M. (2018). Intervision: Supervision for Peer-supported Open Dialogue Teams. *Unpublished manuscript*.
- Holmesland, A., Seikkula, J., & Hopfenbeck, M. (2014). Inter-agency work in Open Dialogue: the significance of listening and authenticity. *Journal of Interprofessional care*, 28(5), 433-439.
- Jorm, A. (2012). Mental health literacy: Empowering the community to take action for better mental health. *American Psychologist*, 67(3), 231-243.
- Kaufman, E., McDonnell, M., Cristofalo, M., & Ries, R. (2012). Exploring barriers to primary care for patients with severe mental illness: frontline patient and provider accounts. *Issues in Mental Health Nursing*, 33(3), 172-180.
- Laugharne, R., Priebe, S., McCabe, R., Garland, N., & Clifford, D. (2012). Trust, choice and power in mental health care: Experiences of patients with psychosis. *International Journal of Social Psychiatry*, 58(5), 496-504.
- Lester, H., Glasby, J., & Tylee, A. (2004). Integrated primary mental health care: threat or opportunity in the new NHS?. *British Journal of General Practice*, 54(501), 285-291.
- Mental Health Taskforce (2016). *The five year forward view for mental health*.

England: Author.

- Olson, M., Seikkula, J., & Ziedonis, D. (2014). *The key elements of dialogic practice in Open Dialogue: Fidelity criteria*. The University Massachusetts Medical School, Worcester, MA.
- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P., Cooper, J., Eaton, J., Herrman, H., Herzallah, M., Huang, Y., Jordans, M., Kleinman, A., Medina-Mora, M., Morgan, E., Niaz, U., Omigbodun, O., Prince, M., Rahman, A., Saraceno, B., Sarkar, B., De Silva, M., Singh, I., Stein, D., Sunkel, C., & Unützer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553-1598.
- Penttilä, M., Jääskeläinen, E., Hirvonen, N., Isohanni, M., & Miettunen, J. (2014). Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis. *The British Journal of Psychiatry*, 205(2), 88-94.
- Ramanuj, P., & Pincus, H. (2019). Collaborative care: enough of the why; what about the how?. *The British Journal of Psychiatry*, 1-4.
- Roth, A., & Pilling, S. (2008). *A competence framework for the supervision of psychological therapies*. University College London, United Kingdom.
- Sharma, R., Sharma, S., & Pradhan, S. (2018). Assessing Caregiver Burden in Caregivers of Patients with Schizophrenia and Bipolar Affective Disorder in Kathmandu Medical College. *Journal of Nepal Health Research Council*, 15(3), 258-263.
- Shay, L., & Lafata, J. (2015). Where is the evidence? A systematic review of shared decision making and patient outcomes. *Medical Decision Making*, 35(1), 114-131.
- Tingle, J. (2019). The urgent need to make NHS mental health care safer. *British Journal of Nursing*, 28(6), 400-401.
- Walker, S. (2003). Interprofessional work in child and adolescent mental health services. *Emotional and Behavioural Difficulties*, 8(3), 189-204.
- Wagner, E., Austin, B., Von Korff, M. (1996). Organizing care for patients with chronic illness. *Milbank Quarterly*, 74, 511-44.
- Whitlock, J., Lloyd-Richardson, E., Fisseha, F., & Bates, T. (2018). Parental Secondary Stress: The Often Hidden Consequences of Nonsuicidal Self

CoMFideS Checklist

Service Name and Trust: _____

Date Submitted: ___/___/___

This checklist is designed to help in the completion of the CoMFideS. It is organised into five areas, which relate to the structure of the scale and the way in which the data will be collected. This data will be reviewed to help complete the scale and will be supplemented by questions for managers and staff of the service. The main areas to be covered are set out in column 1 and possible sources of data are suggested in column 2. Please supply any or all information as this will allow for the best assessment of your service. Please list the information you supply in column 3. Do not worry if any items are duplicated on the list, this is still helpful in completing the assessment. Please attach a copy of the relevant documentation to the checklist. The checklist and documentation may be submitted electronically or in a hard copy.

Area	Possible sources of information	Documents supplied (please attach to the checklist)
<p>1. Service information</p> <ul style="list-style-type: none"> • Operation • Staffing • Caseload • Supervision • Organizational structure 	<ul style="list-style-type: none"> • Operational policy • Referral and service protocols • Annual reports • Yearly reports 	<p>1.</p> <p>2.</p> <p>3.</p> <p>4.</p> <p>5.</p> <p>6.</p>
<p>2. Service structure and culture</p> <ul style="list-style-type: none"> • Service ethos and comprehensiveness • Staff training • Supervision • Staff roles • Service capacity • Routine Outcome Monitoring • Service user involvement (co-production) 	<ul style="list-style-type: none"> • Operational policy • Job descriptions • Staff training policy • Supervision policy • Referral and service protocols • Audit reports • Annual reports • Feedback mechanisms and reports 	<p>1.</p> <p>2.</p> <p>3.</p> <p>4.</p> <p>5.</p> <p>6.</p>
<p>3. Access and engagement</p> <ul style="list-style-type: none"> • Access to the service • Providing information • Prompt action 	<ul style="list-style-type: none"> • Operational policy • Referral and service protocols • Structure of assessment (and available tools) 	<p>1.</p> <p>2.</p> <p>3.</p> <p>4.</p>

<ul style="list-style-type: none"> • Identifying support systems • Flexibility of response • Assertive engagement 	<ul style="list-style-type: none"> • Risk management and safeguarding policies and procedures • Policy and procedure for assertive engagement • Audit reports/Service evaluations and quality reports • Annual reports 	<p>5.</p> <p>6.</p>
<p>4. <i>Delivery of care</i></p> <ul style="list-style-type: none"> • Continuity of care • Establishing clinical meetings • Collaborative decision making • Information sharing and communication • Service user involvement in delivery of care • Case coordination 	<ul style="list-style-type: none"> • Description of interventions provided • Team policies, meeting structures and minutes • Referral and service protocols • Information sharing protocols • Case coordination policy and procedure(s) • Risk management and safeguarding policies • Shared decision making policy and materials • Audit reports/Service evaluations and quality reports 	<p>1.</p> <p>2.</p> <p>3.</p> <p>4.</p> <p>5.</p> <p>6.</p>
<p>5. <i>Community linkage and support</i></p> <ul style="list-style-type: none"> • Service linkage • Community links (Practitioner level) • Community links (Support system) • Caregiver involvement and support • Discharge and aftercare 	<ul style="list-style-type: none"> • Operational policy • Referral and service protocols • Audit reports/Service evaluations and quality reports • Fliers and information handouts for service users and referrers • Caregiver support resources and support policies • Discharge plans and procedures • Aftercare policies and protocols • Annual reports 	<p>1.</p> <p>2.</p> <p>3.</p> <p>4.</p> <p>5.</p> <p>6.</p>

CoMFideS Interview Schedule (Service Managers)

Service information

1. *Organizational structure map*: Please describe the structure of your service including points of access and relevant pathways in and out of the service (Draw a map on Item 5)

Service structure and culture

2. *Service ethos and comprehensiveness*: Could you please describe the model of care you provide as a service? Do you consider there is a coherence in the service ethos?
3. *Staff training*: What sort of training do you expect to see on new staff members? Do you provide additional and/or specialized training for new staff members? (Ask for examples)
4. *Supervision*: What sort of supervision arrangements do you have in your service? How frequent? Does it follow a specific format? Can you describe a typical supervision session?
5. *Staff roles*: What is the clinical-to-support staff ratio? Would you say that your staff is aware of their roles and responsibilities? What do you do to avoid clinical time waste (e.g. task repetition, double-bookings, or contradictory advice to service users, staff and other services)?
6. *Service capacity*: What is the average staff-patient ratio of your service? What is the average caseload for the different roles? What are the different professional, theoretical, and/or technical resources available in your service (e.g. psychiatrists, psychologists, nurses, occupational therapists, social workers, support workers, peer-support workers, etc.)?
7. *Routine outcome measurement*: How does your service seek feedback about the

quality of your practices, procedures, and outcomes? Do you use any PROMS, PREMS What sort of stakeholders do you address (e.g. service users, carers, staff, other services, commissioners, etc.)? Can you give some examples on how you have addressed any concerns? Do you have any quality benchmarks?

8. *Safety*: How does your service ensure your service-user's safety? How does your service ensure your staff's safety? What is the general feedback of your team members regarding their own safety? Have there been any concerns?
9. *Service-user involvement in co-production*: How do service-users get involved in the development, planning, evaluation, and/or assessment of the service as a whole? Provide some examples.

Access and engagement

10. *Access to the service*: How do you try to ensure that the service stays adherent to the operational policy? What type of information do you provide service users about the service? How do you communicate it?
11. *Providing information*: What do external/support agencies know about your service and working ethos? If inaccurate, what has been done to inform them? Please describe the contexts where you last 3-5 cases took place. What sort of information do you provide clients and referrers to inform them about the services and models of care available in your service?
12. *Identification of support systems*: How do you identify and note/register your service users' social support systems/networks? Is this standard practice? Do you usually invite them to clinical meetings?
13. *Prompt action*: What is the average waiting time between referral to service and intake? In case of not meeting the policy time limits, how have you tried to reduce the time-frame?

14. *Flexibility of response*: Can you name all different interventions provided in your service? When your service does not provide a needed intervention (i.e. pharmacological, social, psychological, psychoeducation, peer-support interventions), how do you support access to them (Examples)? Please describe the last two/three cases where you had to collaborate or coordinate with crisis and/or community care services.
15. *Assertive engagement*: Additional to face-to-face sessions, what other means do you have to engage with service users (e.g. transportation, tele-conferences, and home visit arrangements)? Can you describe an example of a service user deemed “difficult-to-engage” and the steps taken to engage with them? Who do you commonly try to engage into a service user’s care?

Delivery of care

16. *Continuity of care*: Is the principle of “continuity of care” a core part of your service’s standard practice? Please describe how this is put into practice. What happens when a key staff member is absent or turnover occurs? What about when a service user is referred back into the service after discharge? What about external service providers, are they also included in network meetings?
17. *Establishing clinical meetings*: Can you describe how network meetings take place? Are care coordination/network meetings the standard way of working? Who is in charge of convening these sessions? Who decides their frequency and timing? Where do they take place? What is the proportion of meetings taking place outside the main service building?
18. *Collaborative decision making*: What is the service’s stance towards treatment plans (e.g. rigid, flexible, existent, non-existent, etc.)? Does the service adapt initial treatment plans along the treatment pathway? Provide examples from the

most recent cases where this has been the case. How do you ensure a sense of safety in your service users? Who is in charge of making clinical decisions (especially around risk)?

19. *Information-sharing and communication*: How are patient records/notes and letters created in meetings? Who is included in the correspondence? How is information communicated or relayed to members who could not attend a meeting?
20. *Service-user involvement in delivery of care*: How do service-users get involved in the provision of care (e.g. volunteering, peer support, advocacy, etc.)? Provide some examples.
21. *Coordination of care*: Which systems are put in place to monitor and manage care coordination? How are external providers included in care coordination? Provide examples

Community linkage and support

22. *Service linkage*: How well embedded do you consider your service to be? Do you consider it has enough and efficient links to external agencies or service providers? How does your service attempt to improve its linkage to the local community?
23. *Community links (Practitioner level)*: How often do you work alongside external agencies or third-parties? How are they included into network meetings? How does follow-up and liaison take place in the service? What sort of additional support does the service provide to service users (e.g. basic living needs, benefits and debts, urgent legal/social problems)?
24. *Community links (Support system)*: How does the service take action to help the service users liaise with the wider network? Please describe how inter-agency

work has taken place in the latest 2-3 cases.

25. *Caregiver involvement and support*: How does the service involve a service user's family/caregivers into care? How are their needs considered and supported? How are these decisions shared with service-users? What sort of information do you provide regarding support services (e.g. carers groups, welfare advice, child support)?

26. *Discharge and aftercare*: How are end-of-care and/or transitions discussed and agreed upon? Please describe your last 2-3 discharge meetings and who was involved in them.

Open Dialogue addendum

27. *Transparency*: Do all discussions about the service-user and their network occur with the service-user present? How is the culture of 'nothing about them, without them' enacted in the service (i.e. neither the service-users nor members of their network are talked about when they are not present)?

28. *Self-disclosure*: How are team members expected to share their own lived experiences (self-disclose)? Are there differences in how this is done in intervision and network meetings? Provide some examples.

29. *Intervision frequency*: How frequent are intervision sessions in your team?

30. *Intervision content and structure*: Can you talk us through the structure of intervision in your team? What is the main focus of intervision sessions in your team? Are team members expected to share personal reflections in pairs or groups? How are these reflections shared with the rest of the team members? How does intervision help the team stick to adherence to the key principles of Open Dialogue? How do you share content from the actual cases? Provide some examples

31. *Team self-work*: Are team members encouraged to maintain a regular personal self-work practice (e.g. mindfulness, psychotherapy, meditation, yoga, personal diary, etc.)? Which ones and how often? Are team members encouraged to maintain a daily self-work practice (e.g. mindfulness, psychotherapy, meditation, yoga, personal diary, etc.)?
32. *Open Dialogue Training*: Have all team members completed, or are undergoing a recognized Open Dialogue training?
33. *Open Dialogue Continuing Professional Development (CPD)*: Is there an annual CPD day organised and led by recognised Open Dialogue trainers? How are team members encouraged to attend?

CoMFideS Interview Schedule (Practitioners)

Service structure and culture

1. *Service ethos and comprehensiveness:* Could you please describe the model of care you provide as a service? Do you consider there is a coherence in the service ethos?
2. *Staff training:* What sort of training did you have prior to joining the service? Did the service provide additional and/or specialized training? (Ask for examples)
3. *Supervision:* What sort of supervision arrangements do you have in your service? How frequent? Does it follow a specific format? Can you describe a typical supervision session?
4. *Staff roles:* Would you say there is a balance in the clinical-to-support staff ratio? Would you say that everyone in your service is aware of their roles and responsibilities? What do you do to avoid clinical time waste (e.g. task repetition, double-bookings, or contradictory advice to service-users, staff and other services)?
5. *Service capacity:* What is the average staff-patient ratio of your service? What is your average caseload? Do you consider it appropriate to provide good quality service? What are the different professional, theoretical, and/or technical resources available in your service (e.g. psychiatrists, psychologists, nurses, occupational therapists, social workers, support workers, peer-support workers, etc.)?
6. *Routine outcome measurement:* How does your service seek feedback about the quality of your practices, procedures, and outcomes? Do you use any PROMS, PREMS What sort of stakeholders do you address (e.g. service users, carers, staff, other services, commissioners, etc.)? Can you give some examples on how

you have addressed any concerns? Do you have any quality benchmarks?

7. *Safety*: How does your service ensure your service-user's safety? How does your service ensure your staff's safety? What is the general feedback of your team members regarding their own safety? Have there been any concerns?
8. *Service-user involvement – Co-production*: How do service-users get involved in the development, planning, evaluation, and/or assessment of the service as a whole? Provide some examples.

Access and engagement

9. *Access to service*: How does your service help you stay adherent to the operational policy? What type of information do you provide service users about the service? How do you communicate it?
10. *Providing information*: How do you make sure that external referrers know about your service's remit and inclusion/exclusion criteria? What sort of information do you provide clients and referrers to inform them about the services and models of care available in your service?
11. *Identification of support systems*: How do you identify and note/register your service users' social support systems/networks? Is this standard practice? Do you usually invite them to clinical meetings?
12. *Prompt action*: What is the average waiting time between referral to service and intake? In case of not meeting the policy time limits, how has the service tried to reduce the time-frame?
13. *Flexibility of response*: Can you name all different interventions provided in your service? When your service does not provide a needed intervention (i.e. pharmacological, social, psychological, psychoeducation, peer-support interventions), how does the service support access to them (Examples)? Please

describe the last two/three cases where you had to collaborate or coordinate with crisis and/or community care services.

14. *Assertive engagement*: Additional to face-to-face sessions, what other means does your service have to engage with service-users (e.g. transportation, tele-conferences, and home visit arrangements)? Can you describe an example of a service-user deemed “difficult-to-engage” and the steps taken to engage with them? Who do you commonly try to engage into a service-user’s care?

Delivery of care

15. *Continuity of care*: Is the principle of “continuity of care” a core part of your service’s standard practice? Please describe how this is put into practice. What happens when a key staff member is absent or turnover occurs? What about when a service-user is referred back into the service after discharge? What about external service providers, are they also included in network meetings?
16. *Establishing clinical meetings*: Can you describe how network meetings take place? Are care coordination/network meetings the standard way of working? Who is in charge of convening these sessions? Who decides their frequency and timing? Where do they take place? What is the proportion of meetings taking place outside the main service building?
17. *Collaborative decision making*: What is the service’s stance towards treatment plans (e.g. rigid, flexible, existent, non-existent, etc.)? Does the service adapt initial treatment plans along the treatment pathway? Provide examples from the most recent cases where this has been the case. How do you ensure a sense of safety in your service users? Who is in charge of making clinical decisions (especially around risk)?
18. *Information-sharing and communication*: How are patient records/notes and

letters created in meetings? Who is included in the correspondence? How is information communicated or relayed to members who could not attend a meeting?

19. *Service-user involvement – Delivery*: How do service-users get involved in the provision of care (e.g. volunteering, peer support, advocacy, etc.)? Provide some examples.

20. *Coordination of care*: Which systems are put in place to monitor and manage care coordination? How are external providers included in care coordination? Provide examples

Community linkage and support

21. *Service linkage*: How well embedded do you consider your service to be? Do you consider it has enough and efficient links to external agencies or service providers? How does your service attempt to improve its linkage to the local community?

22. *Community links (Practitioner level)*: How often do you work alongside external agencies or third-parties? How are they included into network meetings? How does follow-up and liaison take place in the service? What sort of additional support does the service provide to service users (e.g. basic living needs, benefits and debts, urgent legal/social problems)?

23. *Community links (Support system)*: How does the service take action to help the service users liaise with the wider network? Please describe how inter-agency work has taken place in the latest 2-3 cases.

24. *Caregiver involvement and support*: How does the service involve a service-user's family or caregivers into care? How are their needs considered and supported? How are these decisions shared with service-users? What sort of

information do you provide regarding support services (e.g. carers groups, welfare advice, child support)?

25. *Discharge and aftercare*: How are end-of-care and/or transitions discussed and agreed upon? Please describe your last 2-3 discharge meetings and who was involved in them.

Open Dialogue addendum

26. *Transparency*: Do all discussions about the service-user and their network occur with the service-user present? How is the culture of ‘nothing about them, without them’ enacted in the service (i.e. neither the service-user nor members of their network are talked about when they are not present)?

27. *Self-disclosure*: How are you expected to share your own lived experiences (self-disclose)? Are there differences of how this is done in intervision and network meetings? Provide some examples.

28. *Intervision frequency*: How frequent are intervision (or group supervision) sessions in your team?

29. *Intervision content and structure*: Can you tell us what is the main focus of intervision sessions in your team? How does intervision help the team stick to adherence to the key principles of Open Dialogue? How do you share content from the actual cases? Provide some examples Can you talk us through the intervision process in your team? Are you expected to share personal reflections in pairs or groups? How are these reflections shared with the rest of the team members? Provide some examples.

30. *Team based self-work*: Is there an on-going and regular programme of self-work within the team (e.g. family of origin, genogram, or other related)? Which ones and how often? Is such work engaged in by all team members? How are you

encouraged? Are there any self-work retreats you are encouraged to attend (prompt for examples)? Is there training or study leave allowance? How often are you encouraged to attend? Are you and your team members encouraged to maintain a regular personal self-work practice (e.g. mindfulness, psychotherapy, meditation, yoga, personal diary, etc.)? Which ones and how often?

31. *Open Dialogue Training*: Have all team members completed, or are undergoing a recognized Open Dialogue training?

32. *Open Dialogue Continuing Professional Development (CPD)*: Is there an annual CPD day organised and led by recognised Open Dialogue trainers? How are team members encouraged to attend? Can you tell us about the last one you attended?

Appendix K

CoMFideS score distributions

Appendix K. CoMFideS score distributions.

CoMFideS item	Variance	Skewness	Kurtosis	Range
<i>Service structure and culture</i>	0.232	-0.38	-0.89	2.33 - 4.00
SSC1. Service ethos and comprehensiveness	0.688	-0.16	-1.53	2.00 - 4.00
SSC2. Staff training	0.216	-0.98	-1.15	3.00 - 4.00
SSC3. Supervision	0.145	-1.91	1.79	3.00 - 4.00
SSC4. Staff roles	0.428	-0.08	-0.42	2.00 - 4.00
SSC5. Staff capacity	0.346	0.05	-0.11	2.00 - 4.00
SSC6. Routine outcome monitoring	0.601	0.46	0.30	1.00 - 4.00
SSC7. Safety	0.580	-0.67	-0.90	2.00 - 4.00
SSC8. Service-user involvement in co-production	0.984	0.61	-0.48	1.00 - 4.00
<i>Access and engagement</i>	0.232	-0.38	-0.89	2.33 - 4.00
AE1. Access to the service	0.810	-0.13	-1.00	1.00 - 4.00
AE2. Providing information	0.601	0.30	-0.02	1.00 - 4.00
AE3. Prompt action	0.650	-0.60	-1.17	2.00 - 4.00
AE4. Service-user's support systems	0.955	-1.07	0.20	1.00 - 4.00
AE5. Flexibility of response	0.114	-2.42	4.21	3.00 - 4.00
AE6. Assertive engagement	0.514	-1.47	0.80	2.00 - 4.00
<i>Delivery of care</i>	0.382	0.05	-1.05	2.00 - 4.00
DC1. Continuity of care	0.580	-0.94	1.54	1.00 - 4.00
DC2. Establishing clinical meetings	0.580	-0.67	-0.90	2.00 - 4.00
DC3. Collaborative decision making	0.650	-0.60	-1.17	2.00 - 4.00
DC4. Information sharing and communication	0.650	0.60	-1.17	2.00 - 4.00
DC5. Service-user involvement in the delivery of care	1.288	0.34	-1.26	1.00 - 4.00
DC6. Coordination of care	0.636	-0.80	0.74	1.00 - 4.00
<i>Community linkage and support</i>	0.306	-0.50	-0.92	2.20 - 4.00
CLS1. Service linkage	0.601	-0.15	-1.26	2.00 - 4.00
CLS2. Community links (Practitioner level)	0.435	-0.99	0.00	2.00 - 4.00
CLS3. Community links (Support system)	0.928	-0.68	-0.88	1.00 - 4.00
CLS4. Caregiver involvement and support	0.688	-1.46	1.81	1.00 - 4.00
CLS5. Discharge and aftercare	0.630	-0.50	-1.20	2.00 - 4.00
<i>Open Dialogue addendum</i>	1.334	0.09	-1.84	1.00 - 4.00
OD1. Transparency	1.172	0.36	-1.42	1.00 - 4.00
OD2. Self-disclosure	1.297	0.22	-1.34	1.00 - 4.00
OD3. Intervision frequency	1.955	-0.26	-1.92	1.00 - 4.00
OD4. Intervision content and structure	1.993	0.08	-2.00	1.00 - 4.00
OD5. Team self-work	1.275	0.44	-1.20	1.00 - 4.00
OD6. OD Training	1.993	0.08	-2.00	1.00 - 4.00
OD7. OD Continuing Professional Development	2.254	0.13	-2.13	1.00 - 4.00

Appendix L

Mean CoMFideS scores across models

Appendix L. Mean CoMFideS scores.

C-FiT item	Open Dialogue (n=6)		Care as usual (n=6)	
	Mean (SD)	Range	Mean (SD)	Range
SSC1. Service ethos and comprehensiveness	3.42(0.49)	2.50 - 4.00	2.75(0.69)	2.00 - 4.00
SSC2. Staff training	3.67(0.26)	3.50 - 4.00	3.75(0.27)	3.50 - 4.00
SSC3. Supervision	4.00(0.00)	4.00 - 4.00	3.67(0.41)	3.00 - 4.00
SSC4. Staff roles	3.08(0.20)	3.00 - 3.50	3.08(0.74)	2.00 - 4.00
SSC5. Staff capacity	2.67(0.41)	2.00 - 3.00	2.92(0.49)	2.50 - 3.50
SSC6. Routine outcome monitoring	2.00(0.84)	1.00 - 3.00	2.17(0.26)	2.00 - 2.50
SSC7. Safety	3.33(0.52)	2.50 - 4.00	3.33(0.68)	2.50 - 4.00
SSC8. Service-user involvement in co-production	2.00(1.05)	1.00 - 3.50	2.25(0.52)	1.50 - 3.00
AE1. Access to the service	2.67(0.75)	1.50 - 3.50	3.08(0.66)	2.00 - 4.00
AE2. Providing information	2.42(0.49)	2.00 - 3.00	2.42(0.49)	2.00 - 3.00
AE3. Prompt action	3.00(0.71)	2.00 - 4.00	3.58(0.58)	2.50 - 4.00
AE4. Identifying support systems	3.75(0.42)	3.00 - 4.00	2.67(0.88)	1.50 - 4.00
AE5. Flexibility of response	2.83(0.26)	3.50 - 4.00	3.92(0.20)	3.50 - 4.00
AE6. Assertive engagement	3.92(0.20)	3.50 - 4.00	3.25(0.88)	2.00 - 4.00
DC1. Continuity of care	3.58(0.49)	3.00 - 4.00	2.75(0.69)	1.50 - 3.50
DC2. Establishing clinical meetings	3.67(0.41)	3.00 - 4.00	3.00(0.55)	2.50 - 4.00
DC3. Collaborative decision making	3.83(0.26)	3.50 - 4.00	2.75(0.52)	2.00 - 3.50
DC4. Information sharing and communication	3.08(0.92)	2.00 - 4.00	2.33(0.26)	2.00 - 2.50
DC5. Service-user involvement in the delivery of care	2.75(0.94)	1.50 - 4.00	2.00(0.95)	1.00 - 3.50
DC6. Case coordination	3.17(0.68)	2.00 - 4.00	3.08(0.58)	2.50 - 4.00
CSL1. Service linkage	3.08(0.66)	2.00 - 3.50	3.08(0.74)	2.50 - 4.00
CSL2. Community links (Practitioner level)	3.58(0.38)	3.00 - 4.00	3.42(0.58)	2.50 - 4.00
CSL3. Community links (Support system)	3.75(0.42)	3.00 - 4.00	2.58(0.49)	2.00 - 3.00
CSL4. Caregiver involvement and support	3.50(0.55)	2.50 - 4.00	3.33(0.75)	2.00 - 4.00
CSL5. Discharge and aftercare	3.42(0.66)	2.50 - 4.00	3.08(0.74)	2.00 - 4.00
OD1. Transparency	3.00(0.55)	2.00 - 3.50	1.08(0.20)	1.00 - 1.50
OD2. Self-disclosure	3.33(0.52)	2.50 - 4.00	1.50(0.45)	1.00 - 2.00
OD3. Intervision frequency	4.00(0.00)	4.00 - 4.00	1.42(0.49)	1.00 - 2.00
OD4. Intervision content and structure	3.75(0.42)	3.00 - 4.00	1.08(0.20)	1.00 - 1.50
OD5. Team self-work	2.50(1.00)	1.00 - 3.50	1.83(1.13)	1.00 - 4.00
OD6. OD Training	3.67(0.41)	3.00 - 4.00	1.17(0.41)	1.00 - 2.00
OD7. OD Continued Professional Development	3.58(0.41)	3.00 - 4.00	1.00(0.00)	1.00 - 1.00