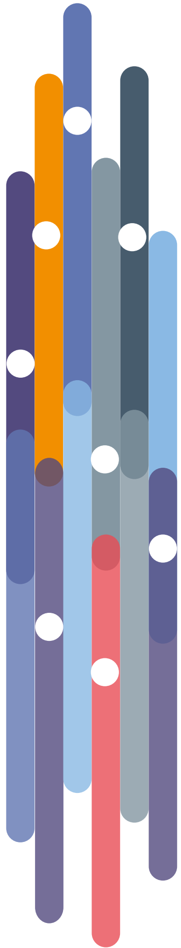


Measuring What Matters

Scoping Review:
The current use of outcome measures by Specialist Parent-Infant Relationship and Infant Mental Health Services

Centre for Early Child Development

Acknowledgements

This study was led by Dr Karen Kinloch and colleagues from the Centre for Early Child Development (CECD).

We would also like to acknowledge the expert advice and support from Karen Bateson and colleagues at the Parent Infant Foundation (PIF). This work would not have been possible without the contributions from practitioners working in specialised parent-infant teams across the UK, who gave their valuable time to complete the survey and participate in interviews sharing their insight and experience which has informed the learning points. The research was also designed in partnership with the Blackpool Parent-Infant Relationship Service Steering Group drawn from a range of disciplines and professions across health, local authority, clinical network, third sector and maternity voices partnership.

This study has received approval from the National Society for the Prevention of Cruelty to Children Research Ethics Committee.

Contents

Acknowledgements 02
Contents 03
Executive Summary 04
Recommendations for practice 06
Useful acronyms 07
Introduction 08

Section 1. Learning from the literature 10
What we did 11
What we found in UK policy 11
What is reported in academic literature? 15
Some conclusions from the literature 28

Section 2. Learning from practitioners 30
What we did 31
Survey responses 31
Learning from the qualitative data 35

Section 3. Conclusions from the study - evidence and practice 49

Section 4. Recommendations 52

References 53

Appendix 1. Literature Search Strategy 58
Appendix 2. Measures used by practitioners 59
2.1 Outcome measures which specialised parent-infant relationship services have used 59
2.2 Measures used by practitioners (survey respondents) 60
Appendix 3. Survey Respondents Job Role and Experience 62
3.1 Job role 62
3.2 Years of experience working with families of children under 5 years 62
3.3 Interventions used with families 63

Executive Summary

Sensitive, responsive, and trusted relationships with a parent are essential to infant mental health and have been shown to positively impact a child across the full range of developmental outcomes (National Scientific Center on the Developing Child, 2012; Barlow et al., 2016; Leach, 2018). Current UK policy initiatives are driving the growth of services to provide support for families to develop these positive parent-infant relationships, including provision of specialised parent-infant relationship teams (Bateson, 2019).

However, identification of appropriate measures for capturing outcomes of parent-infant work remains difficult and this can hinder the development of services (Olander et al., 2021). There have been several systematic reviews of outcome measures for parent-infant services (Coates & Auty, 2019), but there is limited research pertaining to the implementation and acceptability of these for practitioners. This study addresses this evidence gap. Bringing together guidance on measures used to evaluate interventions which address parent-infant relationships by professional bodies, in academic literature and through the voices of practitioners.

The aim of this review is to help guide good practice in evaluation of parent-infant relationship services and to provide practical solutions for future use.

The study highlights the lack of consensus from literature and in practice as to which outcome measures should, and are, being used to assess parent-infant relationship work. Despite the inherent focus on improved attachment as the primary goal of parent-infant relationship work, this was infrequently used as a primary measure in practice, often appropriating proxy measures such as parental mental health for improvements in the relationship. In addition, very few evaluations of services in practice utilised measures of infant mental health, suggestive of a lack of available and suitable mental health measures for this age group (0-2 years).

There were substantive differences between the published guidance on "gold-standard" measures, often associated with academic/laboratory research such as randomised controlled trials (RCTs) and practitioner reported use of measures. Practical constraints, including time requirements, potential burden and

Sensitive, responsive, and trusted relationships

accessibility for families were central for practitioners. A practitioner is likely to view their therapeutic work with the family as their key priority rather than being, as one practitioner described, “the person with the bunch of paperwork” who focuses on evaluating the service.

Nonetheless, practitioners expressed their desire to grow the evidence base for their work, not only for funding purposes, which was raised as an important factor, but to demonstrate for themselves and families the value of the work being done. The ability to “visualise” progress throughout the intervention was a key theme, measures such as Goal Based Outcome measures were often well regarded. It was, however, felt that the need for quantitative measures that could illustrate progress numerically, were more highly valued by commissioners and funders.

Despite the lack of observational measures of the parent-infant relationship reported in use by practitioners, many responses emphasised the need to “see” the relationship, with much clinical practice focused on “seeing what is happening in the room.” Those observational measures which were used overall were well received, but there were concerns over the reliability of ratings. Several practitioners identified that due to time and staffing constraints their preference is to use short questionnaires or surveys to capture an aspect of parent or infant wellbeing from the parent’s perspective. This raises concerns, drawn from the wider discussion, that these may fail to incorporate the expertise of the practitioner in observing the parent-infant relationship.

Recommendations for practice

Five recommendations for practice have been drawn from the learning in the report which can be used to guide the development of outcome measures for parent-infant relationship services.

1

Being Realistic

There is no single measure, or even a set of measures, which can be all things to all people. It became clear during this study there is an ardent desire for a simple, easy to use and universally recommended measure, and that through it we might identify just such a measure. However, we returned multiple times to the conclusion, it is not possible to measure the vast range of potential parent-infant relationship outcomes and, in attempting to do so, we may be doing a disservice to the complexity of the work to support parent-infant relationships.

2

Seeking Clarity

There is a lack of clarity around what the primary outcome should be, and the suitability of measures for outcome measurement as opposed to screening/ assessment. If it can be agreed that the primary outcome for services is an improvement in the parent-infant relationship, then we need clarity as to what aspects of that can be captured in a measure and should be cautious of the use of proxy measures.

3

Capturing Observation

Responses from practitioners showed a real focus on "seeing" the child and the relationship, and how their expertise in this is crucial to the work done with families but is often not captured as part of evaluation. Given the "gold standard" focus on observational measures of attachment, there is value in committing resources to implement and carry out observational measures to see what happens in the parent-infant relationship during, and following, intervention. It should be noted that some interventions have this built-in in an informal manner, for example, Video Interaction Guidance (VIG).

4

Thinking Long-term

To understand the impact on child outcomes, given the complex funding landscape and drivers to deliver evidence in a short period of time, it is necessary to challenge short-term thinking. This extends across giving time for services to do the work and evaluations to capture the impact, but also in implementing measures and then using them consistently.

5

Working Together

A range of stakeholders are involved in the development and delivery of parent-infant relationship work: practitioners, parents, researchers, commissioners, service managers, and national bodies, and it is important these voices are all heard and valued in the identification of measures and development of evaluation. Accepting there will be competing priorities, and that no measure can be all things to all people, is a good starting point from which to build a shared understanding, and to underpin the implementation of the other four learning points of this review.

A note on Language:

When conducting this review, some inconsistencies in common terms used in parent-infant work were identified, with a variety of meanings expressed depending on profession and context. To offer clarity for this review, the terms have been defined as below:

What do we mean by infant?

Traditionally the term "infant" refers to a young child from 0-12 months of age, but for the purposes of this review we are aligning the label of "infant" to include the perinatal period as defined by the National Health Service (NHS) Long Term Plan as the period from conception to a child's second birthday. In some contexts, this is referred to as the first 1001 days (Wave Trust, 2013).

What do we mean by infant mental health?

Infant mental health can be defined as the social and emotional wellbeing and development of children in the earliest years of life. It reflects whether young children have the secure, responsive relationships that they need to thrive (Bateson et al., 2019).

What do we mean by parent?

As the focus of this study was the parent-infant relationship community of practice, we have used the word parent throughout this review when talking about any primary caregiver of an infant. A primary caregiver can include parents, foster carers, grandparents or others that may be formally undertaking this role. The exception to this is where we report on a study on a specific gendered parenting group, i.e., mother/ father or in direct quotes from practitioners' interview and survey responses.

What do we mean by evaluation?

Evaluation in the context of this report means the systematic appraisal of the impacts or outcomes of a service or intervention. This is also referred to as service evaluation in this document.

What do we mean by outcome measure?

The research and practice communities use many labels for the items which are used to assess a client's current health status including: tools, instruments, surveys, questionnaires, and measures. In this review, as the focus is on outcome measurement, we use the term "measure" throughout to encompass any activity which is used to capture the state of the client at a defined point in time (usually before and after intervention). The "outcome" element indicates that the measure is being used to assess whether there is a change following the intervention. This is often framed as whether the intervention is effective or not.

What do we mean by practitioner?

The term practitioner refers to any professional providing support to families where the primary need, and purpose of the intervention, is the resolution of issues with the parent-infant relationship.

Acronym Explanation

CAMHS	Child and Adolescent Mental Health Services
CECD	Centre for Early Child Development
CORC	CAMHS Outcome Research Consortium
CORE	Clinical Outcomes in Routine Evaluation
MORS	Mothers Object Relations Scales
NICE	National Institute for Health and Care Excellence
NHS	National Health Service
PIF	Parent-Infant Foundation
PIP	Parent-Infant Partnership
PRF	Parental Reflective Functioning
RCPsych	Royal College of Psychiatrists
RCTs	Randomised Controlled Trials
VIG	Video Interaction Guidance
VIPP	Video-feedback Intervention to promote Positive Parenting

Introduction

A loving, responsive relationship between infants and their primary caregiver is essential to support children's development physically, socially, emotionally, educationally, and relationally. Most research and practice in early child development emphasises the need to support this primary relationship in the first 1001 days of life to benefit the wellbeing of both parents and infants.

Several initiatives to encourage the development of a good parent-infant relationship have been put in place in the UK in recent years, including antenatal and postnatal parenting programmes, specialist health visiting support, peer support, parent-infant psychotherapy and specialised parent-infant relationship services for those who may need the most intensive support for their relationship.

Health and social care policy in the UK has foregrounded the importance of providing services to support families with infants through the NHS Long Term Plan, the first 1001 days movement, and the Best Start for Life. These initiatives, in practice and policy, are only part of the puzzle to support Infant Mental Health through healthy, responsive relationships. While the Blackpool Better Start partnership worked to develop a new specialised parent-infant relationship service in 2021, it became clear from discussions with commissioners, practitioners and researchers that a crucial element of service delivery was finding ways to measure the impact of specialised parent-infant relationship services.[1]

In 2015, Blackpool was one of five areas in England awarded funding from the National Lottery Community Fund as part of a ten-year strategic investment, A Better Start, with the ambition to develop new approaches to improve early child health and development. Blackpool Better Start is an inclusive partnership approach that includes the National Society for the Prevention of Cruelty to Children, health services, local authority, criminal justice services, and the voluntary sector which collectively lays foundations that enables babies, young children and families to thrive in the context of sensitive and responsive relationships.

The CECD is the research and development hub of the partnership – by using a place-based approach, the partnership is working collaboratively to change the way the community and professionals work to mitigate the impact of early adversity, build resilience and improve outcomes for generations to come. The need for a specialised parent-infant relationship service in Blackpool was included in the vision for Blackpool Better Start from its inception.

[1] The Parent Infant Foundation Toolkit (2019) provides a more detailed rationale of the purpose of measuring outcomes.

Although there is considerable research into the impact of parent-infant relationships on infant mental health and development, during the planning and implementation of this service in Blackpool, the CECD found that researchers and practitioners in this field described a lack of outcome measures which could be used for service evaluation in practice and be reported to commissioners and funders. Without standardised measures, those evaluating the services sometimes struggled to report the difference that the interventions had made to families. When the CECD was in the development stage of the new specialised parent-infant relationship service in Blackpool, we felt it was timely to review the current use of outcome measures used to evaluate this type of intervention, both to inform the service evaluation in Blackpool but also to provide some clarity for others working to evaluate parent-infant relationship interventions.

A brief review of the measures which have been used to evaluate interventions to support the parent-infant relationship and infant mental health, revealed the range of potential ways to measure the impact of this work. This was due, in part, to the many aspects of parent-infant relationships and mental health which could potentially be changed through parent-infant relationship interventions. This initial exploration found no standardised metric of how a "good" parent-infant relationship functions, and although there are several screening tools which can alert practitioners to potential disruption to this relationship, there was no positive equivalent to these.

This study seeks to address this evidence gap, bringing together guidance on measures by professional bodies, a review of measures used to evaluate interventions which address parent-infant relationship, and the voices of practitioners in the field of parent-infant relationships. In writing this review, we aim to amplify the voices of practitioners to help guide good practice in evaluation of parent-infant relationships, and to provide practical solutions for future use.

Crucial to this work is a grounding in the practical use of measures, and the perceptions and experiences of the practitioners who are required to report on the outcomes of services to support parent-infant dyads.

Method

From July to December 2021, researchers at the CECD (supported by the PIF) carried out 3 pieces of connected work into the use of outcome measures by parent-infant relationship/infant mental health services in the UK. The CECD;

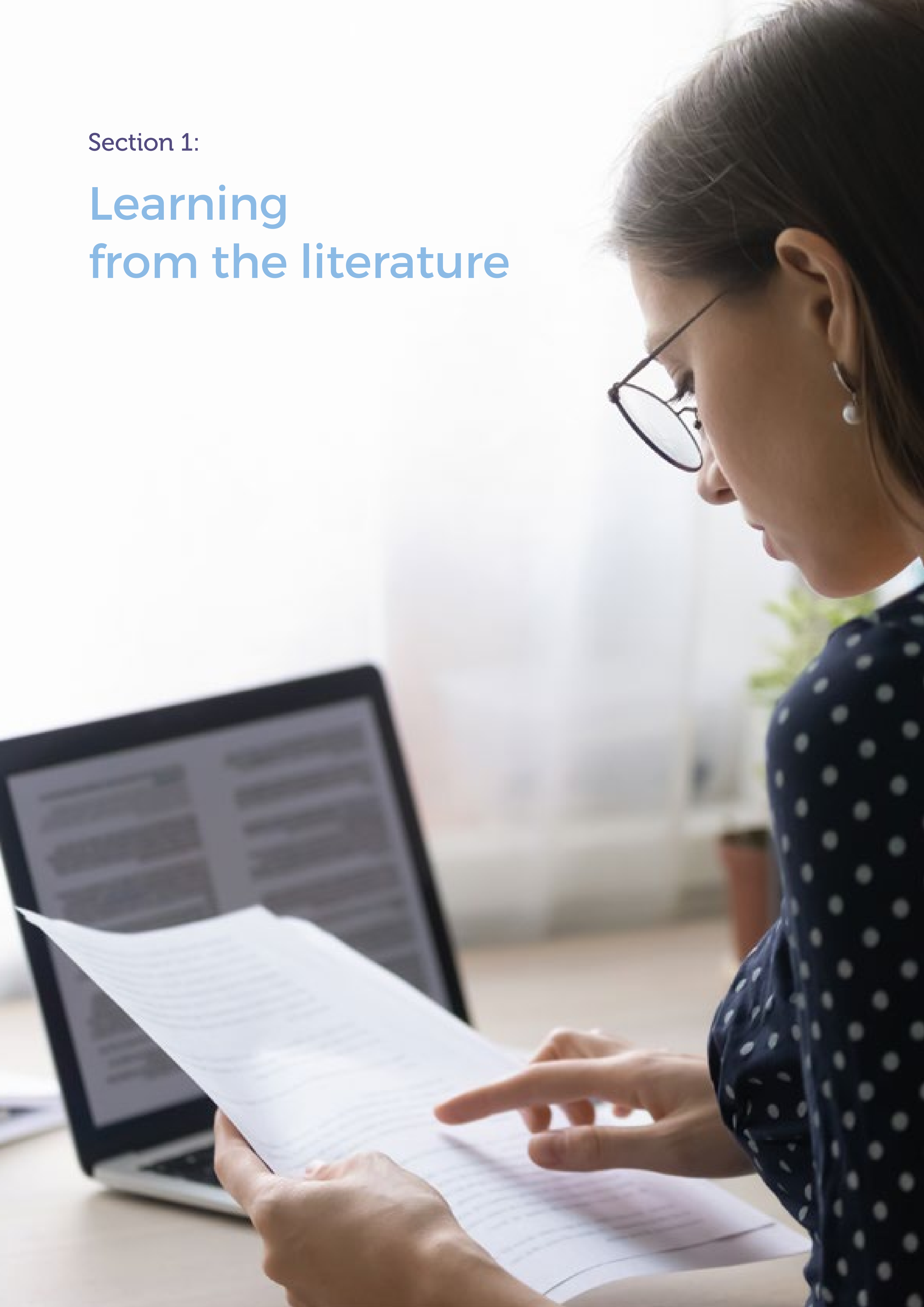
1. Reviewed published literature on the use of outcome measures in parent-infant relationship/infant mental health, this included academic papers, published service evaluations, and policy guidance.

2. Heard from practitioners working in parent-infant relationship/infant mental health services in the UK who completed an online survey about the use of outcome measures in their current practice, and gave their views and experience on the use of outcome measures in this field.

3. Talked to parent-infant relationship/infant mental health practitioners who had completed the survey, to explore in more depth their experiences and views on using outcome measures in their work.

Section 1:

Learning from the literature



What we did

From July to December 2021, researchers at the CECD undertook a mapping exercise of the use and perceptions of outcome measures in UK parent-infant relationship/ infant mental health services. The first stage of this was the identification of the measures which are currently available and recommended by published professional guidance, and academic literature aimed at practitioners working in these fields.

The review aimed to address the following research questions:

1. What type of outcome measures are currently used in evaluation of interventions and services which specifically aim to address disruption in the parent-infant relationship or infant mental health in parent-infant dyads, and
2. What type of outcome measures are recommended for use in parent-infant relationship/ infant mental health interventions and services in current UK policy documents?

The review included three types of publication:

1. Published guidance and policy from National Institute for Health and Care Excellence, the Royal College of Psychiatrists, Institute of Health Visiting, Royal College of Midwives, Child and Adolescent Mental Health Services Outcome Research Consortium, Welsh Government/ Llywodraeth Cymru, PIF, and the Early Intervention Foundation,

2. Published evaluations of specialised parent-infant relationship teams including Attachment, Bonding & Communcation Parent-infant Partnership (ABC PiP) (Belfast), Croydon Best Start P-IP, Leeds Infant Mental Health Service, Little Minds Matter (Bradford), Liverpool PIP, and Together with Baby (Essex), and

3. Peer reviewed journal articles reporting research and evaluation of interventions for parent-infant relationship and infant mental health, including systematic reviews and meta-analyses of outcome measures for such interventions.

It is important to note that this research was conducted in a UK policy context although the academic literature in the review included several international studies where English was the primary language. *For full details of the search strategy used for the literature search please see Appendix 1.*

What we found in UK policy

There is consensus in UK policy on the urgent need for services to support parent-infant relationships, and for measures to evaluate these services. Despite this, there are very few recommendations of specific measures to be used and no standardisation of practice across workforces. Lack of clarity was found on the differences between which measures should be used for screening, and which can be used at repeated time points for outcome evaluation. There was also limited consensus on which aspects of the parent-infant relationship should be used as the primary outcome for evaluation. While there is a focus on attachment as the primary outcome, there are multiple related constructs which are recommended as potential outcomes and the PIF report provides a detailed overview of outcome measures (Bateson et al., 2019; pp. 149-170), a small number of which are used by current services (as seen in Appendix 2).

1: Limited guidance from professional bodies

There is limited published guidance by professional organisations on which tools to use for measuring outcomes in parent-infant relationship work, and there was little consensus on what should be measured.

Parent-Infant Foundation (PIF)

One of the most comprehensive documents on the use of outcome measures currently available is the PIF Toolkit (Bateson et al., 2019), which provides detailed guidance on the implementation of specialised parent-infant relationship services, including suggested outcome measures from both a parent and infant mental health perspective. The Toolkit, however, acknowledges the difficulty in sourcing measures which have ease of access and completion for use in parent-infant relationship work, the tension between the advised use of formal observation for assessment of infant attachment, and the practicality of doing so in day-to-day clinical work.

National Institute for Health and Care Excellence (NICE)

There is limited guidance from the NICE on assessing parent-infant relationships, or potential ways to measure the outcomes of interventions to support this. Both the NICE Clinical Guideline on antenatal and postnatal mental health (NICE, 2020), and NICE Guideline on postnatal care (NICE, 2021) recommend attention is paid to the parent-infant relationship and consider further intervention to improve this relationship if there are problems identified, however there are no recommendations of measures with which to monitor this. An evidence review entitled 'Outcome O' (NICE, 2019) suggests three domains as the most important focus for outcome measures;

1. Maternal feelings towards infant,
2. Maternal-infant interaction, and
3. Insecure attachment.

However, no concrete recommendations are provided of outcome measures which map onto these domains.

Royal College of Psychiatrists (RCPsych)

The RCPsych College Report Framework for Routine Outcome Measures in Perinatal Psychiatry (RCPsych, 2018) highlights the need for outcome measures in perinatal and infant mental health settings, including inpatient settings, and provides some guidance of measures to use. The focus of this framework is understandably on perinatal mental health; however, the role of measures which can be used to evaluate parent-infant relationships (expressed as mother-infant relationships) is still a key aspect of the document.

The RCPsych College Report states that; *"There is very limited evidence base in support of the reliability and validity of measures designed to assess the quality of mother–infant interaction and their suitability for routine clinical practice."*

Institute of Health Visiting and Royal College of Midwives

Similarly, the Institute of Health Visiting (2021) indicate; *"There is currently insufficient evidence to make recommendations about the use of a specific measure of parent-infant relationship".*

The Institute of Health Visiting do not provide any published guidance on use of parent-infant relationship measures. Similarly, the Royal College of Midwives in their updated report on Parental Emotional Wellbeing and Infant Development (2020), highlight the importance of parent-infant relationships and its relationship with perinatal mental health/infant mental health outcomes but no direction is provided to specific interventions or outcome measures.

2: Observational measures of attachment are the gold standard

Policy documents (e.g. RCPsych, 2018; NICE, 2021) consistently recommend the use of observational measures of attachment as the “gold standard” in clinical work, and “attachment” is viewed as the primary outcome of supporting parent-infant relationships in policy literature (i.e. NICE, 2021; PIF, 2019). Furthermore, clinician observational measures of parent-infant interaction are also recommended (RCPsych, 2018).

NICE Quality Standard on Children’s Attachment (NICE, 2016) recommends that assessments of attachment (focussing on children who have experienced care or are on the edge of care) are carried out using validated observational measures based on research by Ainsworth, such as the Strange Situation Procedure (1973).

The RCPsych Framework primarily recommend the use of clinician reported observational measures, e.g. the Parent-Infant Interaction Scale - PIIOS (Svanberg et al., 2013), while also acknowledging the role of parent report measures such as Mothers Object Relations Scales - MORS (Milford & Oates, 2009) for assessing parental thoughts/feelings towards their infant. The Framework indicates that parent report questionnaires are intended as screening tools rather than outcome measures and used to indicate the need for more comprehensive observational assessment.

There appears to be alignment between the recommendations in NICE guidance described above and the RCPsych Framework document on the recommended use of observational measures to assess attachment (also highlighted in the PIF Toolkit as the gold standard for outcome measurement) and to assess parent-infant interaction.

Some exceptions to the preference for observational measures are recommendations from Child and Adolescent Mental Health Services (CAMHS) Outcome Research Consortium (CORC) which include patient report outcome measures such as the Clinical Outcomes in Routine Evaluation (CORE) outcome measure (or short version CORE-10). They also recommend Goal Based Outcome measures and related tools such as the Outcome Star. In contrast with these recommendations by CORC, while the Early Intervention Foundation note that the Outcome Star can be useful therapeutically and for dialogue with clients, they do not consider there is sufficient evidence for use as “*validated evaluation outcome measure*” as it does not meet objectivity criteria needed for formal research evaluation of an intervention.[2]

2 A place for everything and everything in its place: using the Outcomes Stars in combination with validated measures of impact | Early Intervention Foundation (eif.org.uk)

3: Baby blind spot in professional guidance

In NICE Quality Standard [QS133] (2016) there were no recommendations made for measures which can be used with infants under 12 months. Similarly, most measures recommended for use in CAMHS by the CORC are not suitable for work with families of infants under 2 years. An evidence review entitled 'Outcome O' (NICE, 2019) also included no recommendations of intervention work with infants under 12 months, and therefore no outcome measures were recommended for this age group. This dearth of recommendations may contribute to the perception of CAMHS as a service for over 2s, despite the aspiration of services to provide an inclusive service for those aged 0 to 25 years (NHS Long Term Plan, 2022).

"Where available, the gold standard outcome measure for parent-infant work is formal assessment of attachment security." (PIF, 2019)

"clinician-rated observational measures by trained staff are optimal to assess the quality of mother–infant interactions", (RCPsych, 2018)

"The committee agreed that the critical outcomes were the mother's feelings towards the baby when the baby is 12 to 18 months of age, the quality of the mother-baby interactions when the baby is 12 to 18 months of age and the proportion of babies displaying an insecure attachment type when the baby is 12 to 18 months of age." (NICE Outcome O evidence review, 2019)

Box 1. Suggested key recommendations from UK professional bodies on the use of outcome measures for parent-infant relationship work.

What is reported in academic literature?

In the journal articles reviewed, the outcome measures reported for parent-infant relationship work may be categorised into three broad areas;

1. **Parent focused measures,**
2. **Infant focused measures,** and
3. **Measures of the relationship between parent and infant,**
4. **Individual experience and context measures**

Within each category we found there to be a mix of patient reported, clinician rated and clinician observation measures.

1. Parent focused measures

These encompassed parental mental health and perceptions and experiences of parenting, such as self-confidence and competence. The impact of parental mental health and wellbeing on infant developmental outcomes has been well documented, characterised as “adverse effects” (O’Hara et al., 2019) or “contextual stress” (Booth, 2019) in studies of the parent-infant relationship. These “adverse effects” can include parental mental health issues, and lower levels of parenting confidence, competence, and self-efficacy. Multiple studies were identified which focused on improving both parent/infant mental health and the parent-infant relationship, and included adult mental health measures (Barlow et al., 2015; Armstrong et al., 2019; Popp et al., 2019; Rouna et al., 2021; Kristensen et al., 2017; Murray et al., 2016). Several studies/reviews with the primary outcome of sensitive, reflective parenting behaviours also include parental mental health measures due to the potential impact of parental mental health issues on capacity to provide sensitive, responsive parenting (Barlow et al., 2020).

Karitane Parenting Confidence Scale (KPCS)	*Mellow Babies (Raouna et al., 2020), Newborn Behavioural Observation (NBO) (Kristensen et al., 2020) Watch, Wait and Wonder (Cohen et al., 2002) VIG (Adams et al., 2020) and Incredible Years Baby and Toddler (Pontopiddan et al., 2019).*
Parenting Sense of Competence (PSOC)	*Watch Wait and Wonder (Cohen et al., 2002) and Incredible Years (Hutchings et al., 2017).*

Table 1. Most frequently reported measures of parenting confidence/ competence and use in programme evaluation

1.1 Parenting competence/confidence

A lack of parenting self-confidence or competence can be related to both impaired parent-infant interaction and decreased parental mood, therefore many studies of parent-infant relationship interventions include a measure of one of these elements (MacBeth et al., 2015; Kristensen et al. 2020). However, as these are not the identified primary outcome in parent-infant relationship work, attention given to these measures in this review is correspondingly brief and will include only those frequently used in parent-infant relationship intervention studies. For a detailed review of parental self-efficacy measures see Wittkowski et al. (2017) and for a review of parenting confidence scales see Crncec et al. (2010). Two notable scales of parenting competence and confidence most frequently used across a wide range of studies into the efficacy of parent-infant relationship interventions are shown in Table 1. Both measures can be used from birth which may explain their popularity.

1.2 Parental Mental Health Measures

The most common mental health measures in the studies identified assessed mood disorders such as depression, anxiety and stress (Barlow et al., 2015; Henderson et al., 2019; Blower et al., 2019; Blower et al., 2019). More frequently reported in UK studies (Newton et al., 2018) are short self-report measures of mood which are already recommended for general clinical use; the Patient Health Questionnaire (PHQ-9), the Generalised Anxiety Disorder Questionnaire (GAD-7), Depression Anxiety and Stress Scale (DASS-21) and the Hospital Anxiety and Depression Scale (HADS). The short form of the Parenting Stress Index (PSI-SF) has been identified in several reviews as the most common "parent-based factors" measure (McLuckie et al., 2019, Blower et al., 2019). One measure with a broader focus than depression, anxiety or stress is the Brief Symptom Inventory (BSI), a frequently reported measure of global distress in several parent-infant relationship intervention studies (Barlow et al., 2015; Suchman et al., 2017; Fonagy et al., 2016). Almost all parental mental health measures in the literature are patient reported, except for the Structured Clinical Interview - SCID for DSM-IV (Forman et al., 2007; Rayce et al., 2020).

1.3 Parent Experience Measures

There was also widespread use in evaluation literature of individualised measures of the goals, or experiences of parents who engaged with parent-infant relationship interventions. The most common of these was the use of Goal Based Outcome measures which can be used to support and measure progress towards any goal agreed between parent-practitioner, in this context most likely to be related to improvements in the parent-infant relationship (CORC, 2019), and can also be measured using the "Outcome Star" approach. Whatever method is chosen to visualise the data, progress towards a goal is captured on a session-by-session basis.

This category of measures also included service specific Goal Based Outcomes, for example "The Tunnel" in Mellow Parenting (Rouna et al., 2021) - a non-standardised visual analogue scale to capture participants' perceived closeness to their child throughout the programme. Benefits of Goal Based Outcomes include the potential for progress to be non-linear, use of parent report, less linguistic burden, and providing an accessible way to visualise progress.

However, the Early Intervention Foundation do not consider Goal Based Outcomes to be a validated measure as there is no set cut off for "success", making it less useful as an indicator of outcomes achieved. Given this potential challenge, when goal-based data is collected it is worth practitioners and service developers considering what they deem to be a clinically significant change (Wolpert et al., 2019).

Points for consideration in the use of parent-focused measures

Key learning points from the review of parent-focused measures included the lack of consensus regarding suitable measures and the wide range of measures used, and the need to attend to the unique needs and experiences of parents in the perinatal period which are discussed in more detail below.

1a Lack of consensus and significant diversity of measures

Reviews of interventions for parent-infant relationship and for infants reported a wide range of parent focused measures with little or no consensus on their use, despite the substantial body of literature on the importance of parental mental health in relation to infant development. It appears however, from individual evaluations of parent-infant relationship services, that most use short parent report measures of psychological wellbeing such as the PHQ-9 and GAD-7, or CORE-10, often using those recommended by local or national health policy. Evaluation of manualised interventions, such as Incredible Years, Watch Wait and Wonder, and Mellow Babies, frequently use a small selection of measures of parental self-confidence/competence and self-efficacy, but these constructs are not reported as the primary outcome measure in these studies.

1b What are the needs of the perinatal population?

Despite the variety of measures of parental mental health, very few are designed specifically for perinatal mental health except the Edinburgh Postnatal Depression Scale. The scale is widely used as an outcome measure in studies of parent-infant relationship interventions including Mellow Babies, Incredible Years, Enhanced Triple P for Babies, and Video Feedback Approaches (Kristensen et al., 2017; Matthess et al., 2020; Puckering et al., 2010; Henderson et al; 2019) and other non-manualised interventions (Murray et al., 2003; Salomonsson et al., 2021). In the literature there was also a scarcity of measures for fathers' and partners' perinatal mental health, although more measures of parenting confidence and competence are validated for all parents. This may be a result of a focus on maternal mental health in these interventions, although evaluations of family functioning interventions such as For Baby's Sake (Domoney et al., 2019) do include paternal/partner measures.

Stressors relating specifically to parenthood are, in most of the studies reviewed, measured separately to mental health measures as they are contingent on the experience of becoming a parent. However, parental mental health measures need to be utilised considering the experiences of the perinatal period as, for example, Popp et al. (2017) state that "The DASS-21 seemed not valid for new parents. Depressive symptoms such as lack of energy or difficulties to relax are common experiences of new parents". Furthermore, Howard and Khalifeh (2020) in their review of perinatal mental health point out that "infant care itself can generate symptoms that in some studies are attributed to perinatal mental disorders".

Services should also guard against potential use of parental mental health measures as a proxy for improvement in the relationship. While parental mental health can adversely affect outcomes for the infant, including disruption to the parent-infant relationship (Barlow et al., 2020), an improvement in parental mental health alone cannot be shown as proof of efficacy for interventions focused on improving the relationship. Rayce et al. (2020) found no evidence of effect on the parent-child relationship following parenting interventions for mothers with depressive symptoms. Similarly treating only parental depression has not been shown to improve parent-infant relationship, attachment or sensitivity (Forman et al., 2007) and there is a need to consider infant/relationship measures for primary outcomes.

2. Infant focused measures

Infant regulation and attachment problems including emotional, behavioural, eating and sleeping disorders, are prevalent (Skovgaard, 2010; Skovgaard, 2008; Briggs-Gowan et al., 2006) and can predict longer term difficulties in socio-emotional and cognitive development.[3] It is acknowledged that the identification and measurement of these developmental concerns is highly complex and challenging to professionals (Zeneah and Zeneah, 2009). Researchers have identified several risk factors for poor mental health in infants, based on the interaction between the individual child's genetics, temperament and environment (McLuckie et al., 2019).

In the literature we reviewed, the complexity of measuring these factors was evident. Bagner et al. (2012) recommend that further work should be undertaken to explore the relationship between infant temperament, and the relationship with behavioural/ emotional problems. It should also be stressed that infant mental health is most often mediated by environmental factors (Bronfenbrenner, 1977), especially the relationship between infant and caregiver, and these contextual factors should also form part of any robust evaluation.

Currently infant mental health measurement tends to focus on screening and diagnosis (e.g. the Diagnostic Classification Zero-To-Three-Revised Version, [DC 0-3R]), and this impacts the range of suitable tools for gathering outcome measures. However, several systematic reviews and RCTs of measures/interventions for infant mental health were identified (Bagner et al., 2012; Barlow et al., 2015; Fonagy et al., 2016; McLuckie et al., 2019; Szaniecki and Barnes, 2016; Pontoppidan et al., 2017), and we found limited agreement across these about recommended measures. The literature reviewed suggested a wide range of measures are being used to assess infant mental health outcomes and no "gold standard" (Pontopiddan et al., 2017) exists, even across studies that were similar in nature, making it difficult to interpret the impact of interventions on outcomes (McLuckie et al., 2019). The range of measures may be linked to the difficulty in identifying mental health issues in very young children due to the rapid rate of development in this period and the wide range of developmental norms in infants under two years (Pontopidden et al., 2017; Szaniecki and Barnes, 2016).

The included systematic reviews (McLuckie et al., 2019; Pontoppidan et al., 2017; Szaniecki and Barnes, 2016) indicated that all measures reviewed showed acceptable reliability data, although the most psychometrically sound measures which could be completed in a short period of time, and used at repeated time points, were found to be the Ages and Stages Questionnaire: Social and Emotional Second Edition (ASQ: SE-2 [Squires et al., 2015]), and the Brief Infant-Toddler Social and Emotional Assessment (Briggs-Gowan and Carter 2006).

If a longer more detailed measure is required, reviewers found the Child Behaviour Checklist, (CBCL [Achenbach, 2011]) and Infant-Toddler Social and Emotional Assessment to be most valid and appropriate, however the CBCL is limited by the age range it can be used with (1½ - 5 years).

[3] Should be noted that outcome measures have been considered separately to diagnostic tools such as the DC 0-3. The Parent-Infant Relationship Global Assessment Scale (PIRGAS)

Points for consideration in the use of infant focused measures

The key learning points identified in the review of infant focused measures, were the need to consider whether carer report measures adequately capture the experience and needs of the infant, the limited range of measures which are validated for use with infants under 12 months, and the need to include measures which can be used at repeated time points, as opposed to screening or identification tools.

2a Whether to use caregiver reported or observational measures

Many infant focused measures are reliant on caregiver reports of infant behaviours which has both benefits and challenges. Szaniecki and Barnes (2016) note that a major advantage of parent report measures is that they draw on the extensive knowledge parents have about their infant across context and time. Similarly, Pontoppidan et al. (2017) acknowledge the role of observational measures but point out issues with wide variation in development and sensitivity to contextual changes. However, parent reports may be skewed by negative perceptions of the infant due to parental mental health and/or social cultural effects (Peters et al., 2019).

Observational measures usually look at the caregiver-infant dyad relational behaviour, not specifically infant behaviours or temperament; when Bagner et al. (2012) reviewed a range of assessment measures for emotional and behavioural issues in infants, the observational screening was focused on parent-infant interactive behaviours, for example, Parent-Child Early Relational Assessment, and Emotional Availability Scales. While these measures include subscales of infant responsiveness to parental cues, they are more suitable for measuring the quality of dyadic interaction than individual infant temperament (see Section 3 Parent-Infant Relationship focused measures).

2b There are limited measures for infants aged 0-24 months

This literature review identified almost no measures of infant mental health for infants under one year of age; this may be due to the fact it is harder to identify issues in very young children considering the wide range of developmental norms in infants under two (Pontopidden et al., 2017, Szaniecki and Barnes, 2019). Pontoppidan et al. (2017) found a far greater range of measures of social and emotional development in older children. This gap in the literature echoes what we found in the policy review. Those measures which were identified in the review as potentially appropriate for capturing infant mental health, such as Child Behavior Checklist for 1½ - 5-year-olds (CBCL/1.5-5; Achenbach, 1991), was aimed at an age range beyond the limit of what an “infant” is defined as.

The only validated caregiver-report measure found to be suitable for use in the first year is Ages & Stages Questionnaire: Social-Emotional, Second Edition ASQ: SE-2 [Squires et al., 2015], although this is not recommended for use with children with a diagnosed disability (Squires et al., 2009). While the Infant Behaviour Questionnaire: Revised (IBQ: R) is for infants 3-12 months of age there is limited evidence for this measure in practice.

2c How to measure change over time

Measures with fewer items, such as the Brief Infant-Toddler Social and Emotional Assessment and ASQ:SE2, are usually screening rather than outcome measures so may not show sensitivity to change over the course of an intervention. Developers of the ASQ:SE2 (Squires et al., 2015) have also introduced the Social Emotional Assessment/Evaluation Measure for specific use to monitor progress in evaluation of interventions, although there is limited literature on the use of this in practice. A psychometric study has been carried out (Squires et al., 2012) and this measure was included in the review by Pontoppidan et al. (2017).

One other issue with developmental outcome measures is that a short intervention period and lack of follow up could mean that any delayed effects on longer term developmental improvements may not be captured.

3. Measures of the relationship between parent and infant

A central tenet of parent-infant relationship work is that infant mental health development is mediated through the relationship between infant and primary caregiver (Zenneah, 2000; Szanierski and Barnes, 2016). External issues with infant mental health, such as emotional regulation (crying, sleeping, fussing), may be an indication there are also problems with the developing parent-infant relationship. As Pontopiddan et al. (2017) highlight in their review of infant social emotional development measures, it is necessary to also assess this relationship with primary caregivers to gain a full picture of an infant's development.

Although there is consensus in literature and policy on the importance of measuring parent-infant relationships to gauge the effectiveness of parent-infant relationship intervention, the number of complex inter-related constructs creates challenges for practitioners/policy makers to agree on one suitable measure. There are measures recommended in health policy and in research as "gold standard" for formal assessment of attachment security (i.e. PIF Toolkit), however the practicality of completing these in therapeutic practice with often highly vulnerable parents and a range of clinical specialisms, needs further consideration.

The only measure which aimed to measure the parent-infant relationship as a global concept and is used in studies of parent-infant relationship interventions, is the Parent-Infant Relationship Global Assessment Scale (PIRGAS; Zero to Three, 1994). This is a scale for relationship adaptation which is dependent on clinical experience using diagnostic tools for infant development.

Initially, studies of interventions to improve parent-infant relationship were reviewed to ascertain the most common outcome measures of the dyadic relationship (Barlow et al. 2015; Salomonsson, 2014). Following systematic reviews of measures of the parent-infant relationship in current use (Lotzin et al., 2015; Gridley et al., 2019; Trombetta et al., 2015; Wittkowski et al., 2020), and following the guidance of the PIF Toolkit – Chapter 8 (2019), the following constructs were identified as facets of the dyadic relationship which can be operationalised in outcome measures;

- **Attachment or Bonding,**
- **Parent-Infant Interaction (including Parental Sensitivity/Responsiveness),**
- **Parental Reflective Functioning (PRF), and**
- **Parental Representations of the relationship/their infant.**

Outcome measures addressing these constructs were identified for this review, and definitions are given at the start of each section.

3.1 Attachment or Bonding

This is a key outcome of work to improve the parent-infant relationship. As a primary outcome of parent-infant relationship work, there is considerable literature on attachment across a range of studies in comparison to some of the other constructs used to measure the relationship between parents and infants. In a recent construct analysis by Ali et al. (2021), attachment was defined as;

"an affectionate, mutually satisfying relationship between a child and a caregiver that serves the purpose of making the child feel safe, secure, and protected"

while bonding was defined as;

"the emotional bond or tie of affection experienced by the parent towards the infant" (Condon & Corkingdale, 1998)

and crucially can begin antenatally in a way attachment cannot. As the terms attachment and bonding can be (incorrectly) used interchangeably in literature and policy, particularly in antenatal studies (Trombetta et al., 2021; Maas et al., 2012), the construct being measured needs clear definition. Observational measures of attachment in parent-infant relationship can be used to identify infant attachment category (secure, avoidant, disorganised, resistant) or change to attachment security (insecure to secure, stable secure, secure to insecure, stable insecure) (Barlow et al., 2015). However, parent report measures can necessarily only measure parent-to-infant attachment (Trombetta et al., 2021) which would be categorised more appropriately as bonding. Similarly, parent report bonding measures tend to focus on the feelings of the parent towards the infant and their developing relationship. According to Wittkowski's (2020) review of attachment and bonding scales, the *"majority of the measures comprise items that are worded as statements on a Likert-scale that typically enquire how the mother is feeling towards the developing fetus or the newborn."* There is, therefore, some congruence here with measures of parental mental representations of the relationship which we have reviewed separately below.

Several systematic reviews of attachment/bonding measures were identified in this project (Gridley et al., 2019; Perelli et al., 2014; Wittkowski et al., 2020; Trombetta et al., 2021; Tryphonopoulos et al., 2016), of which the first focuses on observational measures of attachment and the last three on parent report. Wittkowski et al. (2020) review of parent report measures, included those measuring either attachment OR bonding, although the majority had attachment as their primary outcome. While this review found that all measures were acceptable for completion time, scoring and readability, the review identified a lack of evidence for robust psychometric properties across all measures studied.

The most frequently cited measure of bonding in reviews (Wittkowski et al., 2020; Perelli et al., 2014) and evaluations of parent-infant relationship interventions (Raouna et al., 2021; Tsivos et al., 2018), is the Postpartum Bonding Questionnaire (PBQ-25, Brockington et al., 2001) and this has the strongest evidence (Wittkowski et al., 2020), which may be due to higher use.

The literature included several parent-report scales which included "attachment" in their title but would be more properly counted as measures of bonding. The Maternal Antenatal Attachment Scale (MAAS; Condon, 1993) and Paternal Antenatal Attachment Scale (PAAS; Condon, 1993) were the most frequently used antenatal scales in studies reviewed by Trombetta (2021), and this finding was supported by review of further literature on parent-infant relationship interventions. However, the Maternal Fetal Attachment Scale (MFAS; Cranley, 1981) and the Paternal-Fetal Attachment Scale (PFAS; Weaver & Cranley, 1983) were found to have the most promising administrative properties of the antenatal attachment scales in Wittkowski et al. (2020).

Despite some promising evidence in previous reviews (Wittkowski et al., 2020), Dunn et al. (2021) found the Maternal Postnatal Attachment Scale (MPAS) (Condon & Corkindale, 1998) was unsuitable for routine clinical use as a screening tool due to poor psychometric robustness.

Two potential benefits of parent reported measures of bonding (although this is still sometimes referred to as attachment) are;

1) the prevalence of measures which are designed to be used throughout the first 1001 days (ante and postnatally), and

2) the prevalence of measures which have been validated for use with both male and female caregivers, giving more options which can be used within a wider range of circumstances.

A selection of these measures is shown in Table 2.

Measure	Antenatal and postnatal scales available	Can be used with caregiver of any gender
Prenatal Attachment Inventory (PAI) (Müller, 1993)	Yes	No
Maternal Attachment Inventory (MAI) (Müller, 1994)	Yes	No
Maternal-Fetal Attachment Scale (MFAS) (Cranley, 1981)	No	Yes
Paternal-Fetal Attachment Scale (PFAS) (Weaver & Cranley, 1983)	No	Yes
Paternal Antenatal Attachment Scale (PAAS) (Condon, 1993)	Yes	Yes
Maternal Antenatal Attachment Scale (MAAS) (Condon, 1993)	Yes	Yes
Paternal Postnatal Attachment Scale (PPAS) (Condon, J., Corkindale, c. and Boyce, P. 2008.)	Yes	Yes
Maternal Postnatal Attachment Scale (MPAS) (Condon & Corkindale 1998)	Yes	Yes
Postpartum Bonding Questionnaire (PBQ) (Brockington, et al. 2006)	No	Yes

Table 2. A selection of parent report measures of bonding and populations for use

The literature included only a limited number of observational measures specifically measuring attachment (Kohlhoff et al., 2020) as opposed to interactions. While there is no doubt that the Ainsworth (1979) Strange Situation Procedure (SSP) is considered the "gold standard" of assessing attachment in young children (Letourneau et al., 2015) as a measure of attachment style, a greater number of observational measures were focused on the relationship mediated through interaction behaviours (see next section) or parental sensitivity. Several studies utilised the Attachment Q-Set (AQS) (Waters & Deane, 1995) to assess levels of attachment security, rather than attachment style. Neither the SSP or the AQS is suitable for use with infants under 11 months and both require intensive observation or video recording of parent-child interaction.

3.2 Parent-Infant Interaction

Even as a subset of parent-infant relationship measures, parent-infant interaction draws on multiple approaches and multiple domains which contribute to the concept of interaction. Interaction measures show considerable cross over with parenting behaviours, such as intrusiveness and responsiveness, and as such, the parent-infant interaction measures include those focused on parental sensitivity, emotional availability and contingent responses (Lotzin, 2015; Mesman and Emmen, 2018).

There are considerable convergences in studies of observational measures for parent-infant interaction (Lotzin, 2015), maternal sensitivity (Mesman and Emmen, 2018) and responsiveness, and to an extent attachment behaviours (Gridley, 2019). As interaction requires assessment of the negotiated dyadic relationship it seems natural that there are more observational measures than parent report.

Lotzin (2015) and Gridley (2019) did not report conclusive evidence for use of one observational measure over another to measure parent-infant interaction in RCTs of interventions, and both reviews indicated the psychometric properties of observational measures have been understudied. Lotzin (2015) also raised the issue that none of the observational measure of interaction had been validated with a paternal population. Our review also found a wide range of observational measures of relational behaviours between parent and infant, based on several different theoretical frameworks and recommendation for use in different settings and age groups, but no consensus on a universal measure of interaction for use in outcome studies.

As different observational measures were developed from difference theoretical basis, the use of these can be dependent on the theoretical underpinning of the intervention being evaluated and the circumstance in which the observation is carried out.

Some examples of this are shown in Table 3.

Several measures of parent-infant interaction using coding of observed behaviour were found to be more suitable as screening or risk assessment tools than to measure changes in interaction over time, although the Parent-Child Early Relational Assessment (Clark, 1985) has been recommended for use in intervention/outcome studies (Tryphonopoulos et al., 2016). The Child-Adult Relationship Experimental Index (CARE-Index; Crittenden, 2001) is also described as a screening tool but has been used to measure outcomes post intervention (Barlow et al., 2007; O'Hara et al., 2019), and is used to code multiple aspects of parental behaviour. The Mother and Baby Interaction Scale (MABISC) was unusual as a patient report measure, which was used for outcome evaluation of parent-infant relationship intervention RCTs (Pontopiddan et al., 2019), however it was also recommended for screening not outcome measurement by the measure developers.

Measure	Theoretical Basis
Coding Interactive Behavior (CIB; Feldman, 1998)	Transactional model of development
Parent Infant Relational Assessment Tool (PIRAT; Broughton, 2010)	Psychotherapeutic principles
Parent–Infant Interaction Observation Scale (PIIOS, Svanberg et al., 2013)	Parental sensitivity and mind-mindedness
Child-Adult Relationship Experimental Index (CARE-Index; Crittenden, 2001)	Dynamic Maturation Model of Attachment
Keys to Interactive Parenting Scale (KIPS, Comfort et al., 2011)	Parenting Strengths and Needs based
American Nursing Child Assessment Satellite Training (NCAST) Parent-Child Interaction (PCI) Scales (Sumner, 1994)	Sensitive and responsive caregiving is predictive of secure child attachment

Table 3. A selection of measures and their theoretical underpinnings

3.3 Parental Sensitivity and Emotional Availability

Both parental (often phrased as maternal) sensitivity and emotional availability are recognised as aspects of a responsive and reciprocal parent-infant relationship, and in the literature, measures of these concepts were frequently used to assess both parent-infant interaction and, more broadly, parent-infant relationships.

Maternal Sensitivity as part of parent-child interaction includes appropriate recognition of and response to infant signals, being described as “a vitally important index of individual differences in the quality of early infant–caregiver interaction” (Meins, 2013). The concept of sensitivity, therefore, refers not just to a set of behaviours but to something much more dynamic and relational drawing on concepts developed by Ainsworth (1974). These include the Ainsworth Sensitivity Scale (ASS; Ainsworth, 1974), Maternal Behaviour Q-Sort (MBQS) (Pederson, 1999), and the Parental Sensitivity Assessment Scale (PSAS; Hoff et al., 2004).

Emotional availability is described as the capacity of parent–infant dyads to share an affective connection and enjoy a mutually fulfilling and healthy relationship (Biringen & Easterbrooks, 2012). The most common measure is the Emotional Availability Scales (Biringen, 2000) used to rate video recorded free play interactions between parent-infant, including both parent and infant dimensions. The literature showed frequent validated clinical use of these scales in systematic reviews and RCTs (Backermans-Kranenberg et al., 2013; Matthes et al., 2019; Sprengler et al., 2021). Despite this, Gridley’s (2019) review of observational measures of parent-child interaction found the evidence for psychometric properties of the Emotional Availability Scales is inconclusive as this was based only on three RCTs.

3.4 Reflective Functioning and Mental Representations of the Relationships

Another key area for measurement of the parent-infant relationship is the parent's mental representation of both their relationship with their infant, and the mental processes of the infant themselves.

Parent's perceptions of their child and relationship have been used in trials of parent-infant relationship interventions, for example the Working Model of the Child Interview (WMCI; Zeanah & Barton, 1989) was used in Circle of Security (Mothander et al., 2020) and Watch Wait and Wonder (Rance, 2012) to assess caregivers' experiences of the infant, their representations of the relationship with their infant, and thoughts about the child's future and the impact of these on the relationship.

The most frequently reported parent report measure of the perceived relationship is the MORS, and the short form of this scale (MORS-SF), which elicits parents' 'representations' of their infant's thoughts, feelings, and intentions towards them. There was reported use of the Mothers Object Relations Scales (MORS) in several studies and there appears to be some evidence for the MORS as an outcome measure (Coster et al., 2015; Bhopal et al., 2022), this may just indicate evidence of use rather than evidence of the MORS as the most appropriate tool for measuring change over time in the parent-infant relationship.

The ability of the parent to reflect on their own and their child's internal feelings and moods in relation to their external behaviour, referred to by Fonagy et al. (1995) as "mentalization" or "reflective functioning", is thought to play an important role in the development of the parent-infant relationship (Slade et al., 2005) and on children's ability to develop mentalisation and emotional regulation processes themselves. Barlow et al. (2020) found that there has been limited use of PRF measures in studies of early years dyadic interventions, despite the evidence base which suggest PRF is an important predictor of infant attachment. In this study we also identified fewer measures of PRF in comparison to other aspects of the parent-infant relationship. The Reflective Functioning Scale (RFS; Fonagy et al., 2002) has been used to successfully measure change in PRF in several studies (Sadler et al., 2013; Fonagy et al., 2016). However, the only patient report measure we identified is the PRF Questionnaire (PRFQ; Luyten et al., 2017) which has been widely validated for research use and is the most frequently used outcome measure for PRF in studies of parent-infant dyads (Barlow et al., 2021). This has not yet, however, been recommended for use in clinical settings by the developers as further research is required in these settings.

There is some evidence to suggest that parent's mentalisation of their infant and their relationship may be successfully measured as a predictive measure of attachment (Barlow et al., 2020), and should be considered as part of an outcome evaluation of parent-infant relationship interventions where either PRF or attachment are the primary outcomes.

Points for consideration for use of parent-infant relationship focused measures

Although there is consensus in the literature and policy on the importance of measuring parent-infant relationships to gauge the effectiveness of parent-infant relationship intervention, the number of constructs which feed into the development of this relationship, and the complex inter-relation between these constructs, creates challenges for practitioners/policy makers to agree on one suitable measure.

3a What, therefore is the primary "relationship" outcome for parent-infant relationship interventions?

Although "attachment" is described as the primary outcome for parent-infant relationship work, studies include a greater number of observation measures of behaviours thought to indicate attachment, such as positive, reciprocal interaction between parent and infant, and parental sensitivity to infant cues. This may, in part, be due to the difficulty in measuring attachment following a short-term intervention, as Bakermans-Kranenburg et al. (2013) found attachment insecurity is more difficult to change through intervention than maternal insensitivity. However, this may be due to a sleeper effect (time lag) in changes to attachment security following improvements in maternal sensitivity. It is therefore worth exploring what best reflects a positive change in the parent-infant relationship and using this to identify the most suitable measure of whether the aims of an intervention have been met.

3b What are the differences in measures between research and clinical practice?

Our review of the literature highlighted the difference in research practice and clinical practice in terms of the number, type, and range of outcome measures in place. Research studies which included at least seven measures at multiple time points, and often used observational measures, were in contrast with evaluations of clinical practice where services usually reported a maximum of four measures, often using parent reported measures.

Parent-report measures were found to be less likely to be included in systematic reviews of interventions, as these include only RCTs which use observation measures, for example, Barlow et al. (2015) review of PIP and Bakermans-Kranenburg et al. (2013) review of maternal sensitivity and attachment. Similarly, RCTs of interventions such as VIG for maternal sensitivity (Barlow et al., 2016) use the CARE-Index (Crittenden, 2001). It is likely that observational measures are more commonly used in RCTs research settings than in clinical practice (O'Hara et al., 2019) as, due to the sheer amount of time and need for laboratory setting and trained coders, they are less feasible for use in therapeutic settings.

As parent-report scales are shorter in length and time to administer, they are practical for use in therapeutic settings where there may not be the time or facilities to administer and code observational measures. However, caution should be exercised to ensure measures are used to measure the primary outcome, not the outcome which is easiest to capture and to avoid measuring one construct and assuming this can act as a proxy for all aspects of the relationship.

There is some evidence to suggest that parent's mentalisation of their infant and their relationship may be successfully measured as a predictive measure of attachment (Barlow et al., 2020), and should be considered as part of an outcome evaluation of parent-infant relationship interventions where either PRF or attachment are the primary outcomes.

4. Individual Experience and Context Measures

In addition to psychometric outcome measures, many studies included an element described here as individual focused measures, those measures such as Service User Feedback, Experience of Service Questionnaire and Goal Based Outcome measures which are dependent on feedback from individual clients throughout the intervention period, and those which capture contextual data relevant to the individual families' circumstances, for example the level of involvement of Children's Services.

4.1 Goal Based Outcome Measures

Goal Based Outcome measures can be used to support any goal agreed between parent-practitioner, however in this context they were most likely to be used to measure improvements in goals which were centred around the parent-infant relationship, for example in parent-infant psychotherapy (CORC, 2019). Goal Based Outcomes can be tracked using different tools, such as the Outcome Star or the goal rating sheet and used to visualise progress. This set of measures also included service specific goal-based measures, for example "the Tunnel" in Mellow Parenting (Rouna et al., 2021) - a visual scale developed by the Mellow Parenting Evaluation team to capture participants' perceived closeness to their child throughout the programme.

Potential benefits of Goal Based Outcomes were found to include less linguistic burden on families, and the individual nature of goals setting can be used across a wide spectrum of therapeutic approaches.

However, the fact that individualised measures inevitably require parent report may indicate the need for an additional measure to capture the voice of the infant. This type of measure was rarely, if ever, used to measure the primary outcome in impact evaluations, perhaps due to a perception of them as subjective in comparison to other types of measure.

4.2 Contextual Information

In the studies of parent-infant relationship interventions reviewed for this review, all data collection included demographic information on the families who engaged, as it is acknowledged that social and ecological (Bronfenbrenner, 1977) risk factors can predispose families to issues with the parent-infant relationship and impact on levels of change which can be attributed to an intervention. This data was sometimes collected using a risk and stressors checklist (Balbernie, 2003) to capture the circumstances surrounding the family. Several contextual areas for consideration were highlighted in the literature including; high/low risk study population (Sleed et al., 2013; Mascheroni and Ionio, 2019), clinical or non-clinical population (Hendserson et al., 2019), the age of the infants engaged in the intervention (McLuckie et al., 2019), and gender differences in parenting responses.

A narrative review by Egeland et al. (2000) of 15 attachment-based interventions, pointed out there are many factors at different ecological levels that may interfere with successful intervention and, therefore, tracking these factors is necessary in evaluating of services.

Some conclusions from the literature

Initial analysis of the literature resulted in grouping of the measures into the three broad categories of child outcome focused (child development or behaviour), parent focused (parental mental health or parenting experiences) and parent-infant relationship focused, the latter of which encompassed many possible constructs for measurement. Within each of these categories various patterns were noted, for example patient reported outcome measures were more likely to be used for parent focused measures, while observational measures were more prevalent for measuring factors relating to the infant or to the relationship.

Overall, the literature review revealed the heterogeneity of measures across all the categories, with limited consensus on which measures to use and with no outstanding evidence for any one measure or set of measures. This lack of consensus may be attributed to the highly complex nature of the parent-infant relationship and the number of contributing factors. In the process of the review, specific gaps in the literature were identified the dearth of measures for infant mental health for infants under 1 year of age and the divergence between approaches to outcome measures in academic research and practice-based evaluation.

No clear consensus

The literature reviewed indicated there is no consensus on either the constructs to be measured, or the outcome measures with which to do this in parent-infant relationship work. Systematic reviews of measures did not provide strong evidence to support any measure across a range of domains and for both parent-report and observational measures (McLuckie et al., 2019; Pontoppidan et al., 2017; Szaniecki and Barnes, 2016; Lotzin et al., 2015; Gridley et al., 2019; Trombetta et al., 2015; Wittkowski et al., 2020). Many of the studies reviewed did not often have aligned primary outcomes (beyond a broad remit to improve the parent-infant relationship), and those which did measure the same primary outcome often utilised different measures for the outcomes.

While observational measures are often seen as the "gold standard" for measuring parent-infant relationship (Gridley, et al., 2013) some authors, including Lotzin et al. (2015), suggest that "If a comprehensive evaluation of the parent–infant interaction is needed, this should ideally include both parent-report and observations - these two types of measure are not interchangeable and are only weakly correlated", while Bagner et al. (2013) state that *"A more comprehensive multi-method evaluation would ideally include both questionnaires and behavioural observations".* One inference which can be drawn from this is that researchers agree a multi method approach is necessary to properly evaluate parent-infant relationship interventions, although just which measures should be used is not yet conclusive.

Measuring complexity

One possible reason for the lack of conclusive evidence for a specific measure, or set of measures, may be the sheer number of inter-related facets which contribute to the parent-infant relationship, and which can be measured for outcome evaluation. In this review, the facets identified included attachment, bonding sensitivity, responsiveness, reflective functioning/mentalisation, parent-child interaction, and mind-mindedness. For each facet there were also a variety of measures which could be used to assess outcomes of parent-infant relationship interventions. This level of complexity, and the relationship between factors, may preclude standardisation of measures or lead to many measures being utilised to “capture” the outcomes of an intervention.

Minding the gaps

Despite the large range of measures, several “gaps” were identified around the use of outcome measures. One of these is the “Baby Blindspot”, a paucity of outcome measures for the mental health of infants under 12 months, especially in comparison with measures focused on the mental health of parents. There are more possibilities for measuring behaviour in infants, however some of these are dependent on the infant having reached certain developmental milestones, i.e., more suitable for use in diagnosis and screening rather than outcome evaluation. The lack of outcome measures in the infant mental health literature may imply that this issue is also found by those providing and evaluating services.

A more subtle gap was identified in the use of outcome measures used in service evaluation and in academic research. As discussed previously, there are many factors contributing the parent-infant relationship and in research studies, particularly RCTs, large numbers of assessments can be included to measure a range of these factors and examine potential associations between them. This was also the case for the type of measure used, in guidance and in the majority of research studies the “gold standard” of observational measurement for attachment as the primary outcome was applied while in contrast this was not reported as an outcome in the evaluation documentation. The gap between the breadth of measures available to research studies and those which are used in practice, should be considered when selecting measures based on their feasibility in research as this may not translate to clinical practice.

Section 2:

Learning from practitioners

What we did

The voice of practitioners who are using outcome measures within the field of parent-infant relationships was central to our rationale for this study, and we hoped that data from practitioners would provide context to the outcome measure use that was evidenced in published literature.

For this phase of the study a mixed methods approach was used:

1. An online survey was circulated to practitioners to ask about their current practice around outcome measures, and
2. A volunteer subsample of those who completed the survey also took part in an interview with a member of the CECD Research and Evaluation Team to discuss, in more depth, their experiences and attitudes towards outcome measurement in their role and service.

Survey

The Research and Evaluation Team at the CECD developed a short online survey to explore which measures practitioners are currently using, and how they feel about the barriers/facilitators of completing outcome measures in relation to parent-infant relationship and infant mental health work. In September 2021, the survey went live for six weeks and was distributed to the PIF mailing list (n=819) to reach practitioners working in parent-infant relationships services, although it should be noted that this email list includes anyone with an interest in the work of the foundation. The survey asked respondents about their professional experience, their use of interventions and outcome measures, and their perceptions of the challenges and facilitators to using these measures in parent-infant relationship and infant mental health work.

Interviews

People who responded to the survey were invited to take part in telephone interview with a researcher from the CECD to explore, in more detail, practitioner experiences of using outcome measures in their service delivery. The data gathered through interviews was crucial to expanding our knowledge of how measures are used in practice, and to contrast this with the findings in the literature. The interviews (n=8) were conducted from September to November 2021, and the main themes of the practitioner responses were identified.

Survey responses

Over a six week period the survey received 47 complete responses from practitioners working in parent-infant relationship and infant mental health services. Of these, almost half (n=22, 47%) currently work in a specialised parent-infant relationship team which is part of the PIF Network. Looking at the response rate, there may have been merit in sharing the survey across social media as it was evident that many practitioners are working in roles which are not part of the PIF Network.

Practitioner location

Responses were received from practitioners working across the four nations although most respondents came from teams working in England, with the largest number from the South East (including London) (n=15) and the North West (n=11) and no responses were received from practitioners in the North East of England. The number of respondents in these areas broadly reflects the distribution of parent-infant relationship teams on the PIF Network map, although work is underway to develop teams in currently underserved areas.

Professional experience and role

Responses were received from practitioners with a range of years' experience of supporting parent-infant relationship and infant mental health (see Appendix 3), with the majority (n=22) having more than 10 years' experience, and in a wide variety of professional roles.

Unsurprisingly, given the target population for the survey, most respondents were from a psychological background (n=14) and this category comprises both clinical, educational and research psychologists. Similarly, psychotherapists (n=13), including adult and child specialisms, were well represented. It was pleasing to note that almost a quarter of respondents were early years and family workers from a range of settings including Children's Centres and Peer Supporters. Allied health and social care professionals, such as Specialist Health Visitors and Social Workers, were less well represented and it would be valuable to hear more from this sector of the workforce (for full list of roles see Appendix 3). Similarly, only a small number of those in commissioning and management roles responded. This will possibly impact on the perspectives given about the role and importance of outcome measures.

To get a picture of the work taking place with families which may be evaluated using outcome measures, respondents were asked to tell us about the interventions they are currently using in their work.

Which interventions are used by practitioners?

When asked which programmes and approaches were part of their suite of interventions, respondents most frequently reported use of Video Feedback Approaches (VIG or VIPP), and psychotherapeutic (triadic and dyadic) or psychoanalytic techniques (n=31 and n=30 respectively). Responses also showed moderately frequent use of manualised, usually evidence based, programmes such as Circle of Security (n=12), Watch Wait and Wonder (n=13), Mellow Parenting (n=6), and Incredible Years (n=11). Evidence based approaches, such as Solihull (n=15) and Brazelton (n=6), were also mentioned by respondents. Several different modes of delivery, such as parent-baby groups (n=15), peer support programmes (n=9), and infant massage (n=8), were also reported by multiple practitioners.

Only nine respondents mentioned antenatal interventions and there were limited numbers of respondents using creative or play approaches although this may be due to the professional demographic of respondents as there were no responses received from practitioners working as art therapists. A small number (n=4) use bespoke programmes and parent only therapy (n=3). Several of these interventions have recommended outcome measures built into the programme delivery (e.g. VIG, Mellow programmes), which may influence the reported use of measures in the survey.

Which outcome measures are used by practitioners?

The information from the initial review of policy and academic literature was used to develop a short-list of measures which may be used by practitioners with families prior to engagement (screening), and pre and post clinical intervention (outcomes). In the survey, practitioners were also asked to list any measures not given in the shortlist which they use in their work. These responses showed the wide range of individual measures used, with a total of 48 unique measures named. It should also be noted that some measures on the initial short list were not used by any of the respondents. A full table of the measures reported in the survey can be found in Appendix 2.

A small group of five measures were reported as used by most respondents, these included the PHQ-9, GAD-7, MORS, SDQ, and Service User Feedback. While there were many other measures reported, these were used by only small numbers of respondents (n=1-4 per measure)

Due to many individual measures being identified, they were grouped into outcome domains (see Appendix 4) based on the outcome the measure was designed to capture, for example measures of maternal anxiety were included in the Parental Mental Health and Wellbeing domain. This categorisation gives a broad view of the types of constructs which practitioners are currently using outcome measures for (the total frequency of these domains is shown in Fig 1).

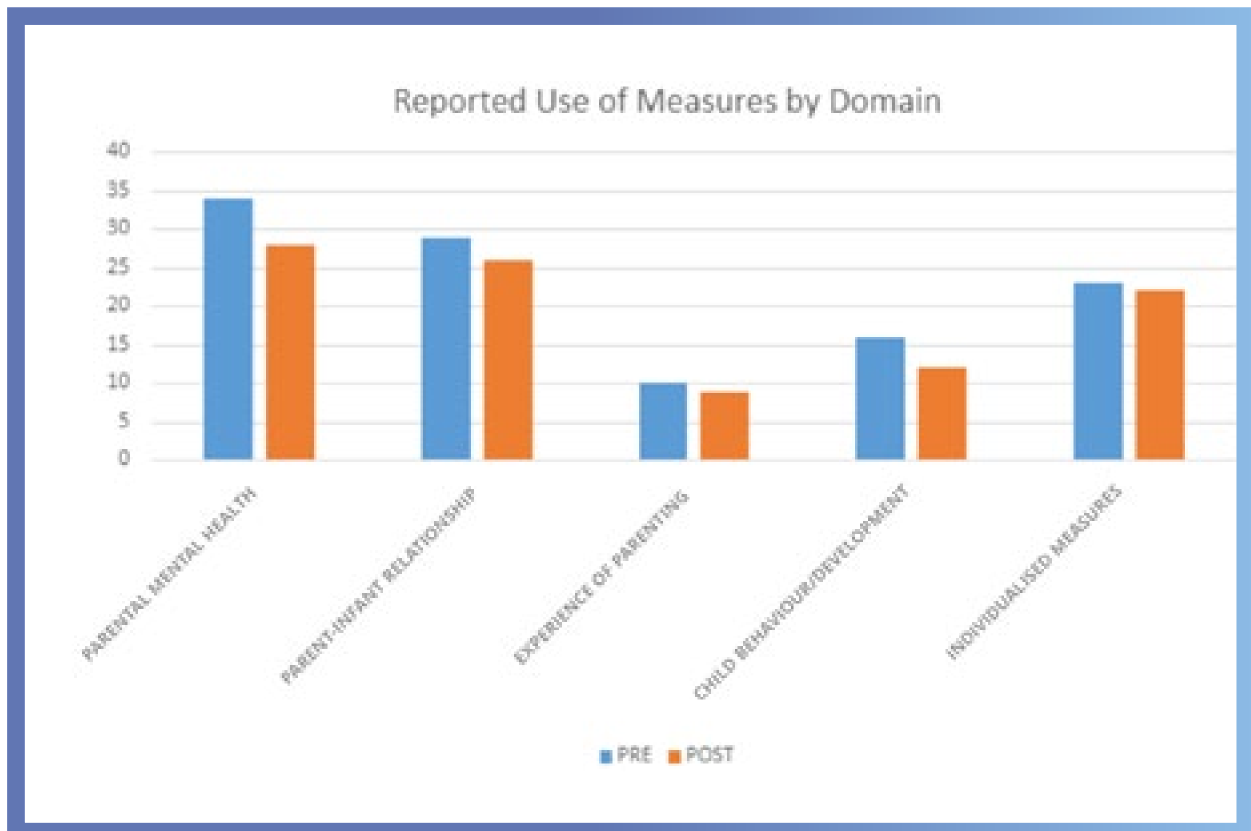


Fig 1. Frequency of reported use of measures by domain (pre and post intervention)

The most frequently reported domain was measures of Parental Mental Health and Wellbeing, closely followed by measures of the Parent-Infant Relationship. A form of individualised experience measures, such as Goal Based Outcomes and Service User Experience, were used by almost half of respondents. Less than a third of respondents reported the use of measures related to child development, or the experience of parenting.

Summary

A small number of frequently used tools for measuring parental mental health and parent-infant relationship, (PHQ-9, GAD7 and MORS), accounted for most reported use in the survey, with a long list of other measures used by a small number of individual respondents. While this suggests some commonalities across services, it also points to a more individualised approach to outcome measurement, echoing the lack of consensus we found in the academic literature. If these responses reflect the pattern of use across Parent Infant Relationship Services in the UK, it would be helpful to undertake an in-depth exploration of the motivation for using the most popular measures as this may be due to standardisation or historic behaviour which is reflected in the qualitative responses to this study.

The reported measures do not reflect the "gold standard" of using observation as described in policy and research (NHS, 2020; Gridley et al., 2019; Condon, 2012). Most measures identified in the survey responses are patient reported short questionnaires, it is plausible that services are tending towards measures which can be completed quickly and easily given the time constraints and burden already on practitioners working in CAMHS, which was identified by Hall (2018) and reiterated in the responses to this survey.

It was slightly surprising, given the focus of the study on those working in parent infant relationship services, that the most frequently reported measures related to the domain of parental mental health rather than the relationship between infant and caregiver. One potential explanation for the parental mental health focus may be the number of respondents who identified working closely with specialist perinatal mental health services, as some areas choose to commission a Perinatal and Parent Infant Mental Health Model which encompasses parent-infant relationship services. Whilst there is evidence of strong links between parental mental health and child outcomes, positive changes in outcomes of parental mental health should be interpreted cautiously and not used as a proxy for measuring the parent-infant relationship, as improvements in parental mental health do not necessarily reflect improvements in the relationship with their infant (Bagner et al., 2019).

The importance of outcome measures

We were interested to find out how important outcome measures are to practitioners working to support parent-infant relationships and infant-mental health. In response to the question "How important is it for your work to collect data using outcome measures?" on a 5-point scale, with a higher score indicating higher importance, the mean was 4.5 with 66% of respondents rating outcome measures as very important.

Learning from the qualitative data

Perceptions of outcome measures and their use in practice

While it was helpful to gather data on the types of outcome measures used within service, of equal importance is the information we obtained on practitioner perceptions of outcome measures and how they are used in work with families. This data was collected using both free text comment boxes in the survey, and in more depth through research interviews with practitioners. The questions asked in the survey and interviews focused on several key areas including the importance and purpose of outcome measures, what barriers and facilitators are involved in using measures, and suggestions for use of measures in the future that would aid good practice.

The qualitative results were thematically analysed using deductive analysis so that the identified themes were based on the prior knowledge of measures gained throughout the study (Braun and Clark, 2013). Across both qualitative datasets there was considerable evidence of tensions between aspects of selecting and using outcome measures in parent-infant work.

> *"Because in a way what we're looking for is this straightforward measure but to measure something that's incredibly complicated." (P03)*

In the qualitative data collected from the survey and interviews, practitioners provided their views on:

- **Identifying and implementing measures,**
- **The purpose of measures for a range of different stakeholders,**
- **The experience of practitioners in using measures, and**
- **Suggestions for improvements which could aid the use of measures in the future, including potential barriers and facilitators to this.**

The themes discussed were identified in both the analysis of the comments in the free text boxes on the survey, and in the in-depth interviews which were carried out with practitioners. While not exhaustive, seven themes reflect the salient points identified for policy and practice around outcome measures;

1 Measuring complexity, simply- looking for the 'Holy Grail'?
2 Quantifying the worth of the work
3 Seeing change - the importance of the visual
4 Measures as a therapeutic tool
5 Feeling the burden – time and space to complete measures
6 Towards shared understanding and practice
7 Thinking long term

1 Measuring complexity, simply - looking for the 'Holy Grail'?

A central issue, particularly for the identification and implementation of outcome measures, is the tension between the varying needs of practitioners, families and commissioners, but also the need to access simple measures of an extremely complex phenomenon which do not place too great a burden on the time of practitioners and families. This theme exemplifies the challenges in identifying a measure which is feasible for use, captures what is considered important and can be completed in a limited time.

Practitioners frequently referred to the ongoing search for measures which are suitable for use in parent-infant work, those who had worked in the field for several years or longer described how they had been looking at ways to measure outcomes for most of this period.

> *"...in that time outcome measures, they've always been problematic, so we have always struggled." (P06)*

> *"So I've been working with the team for six years now, since 2015, and it's just an ongoing discussion and headache really to try and find something." (P05)*

Several respondents also highlighted the complexity of the work which parent-infant relationship service are doing with families, and the difficulties in capturing those complexities in an objective outcome measure which can also be used in clinical practice, particularly in measuring the parent-infant relationship.

> *"I think the challenge is that this is a complicated thing, isn't it, to do that parent/infant observation and to do it with any sense of objectivity." (P04)*

> *"We would like something that measures more the subtlety I think of the relationship, we just don't feel like that we have found anything that really looks at the relationship in the way that we want." (P05)*

While practitioners reported they valued outcome measures, several also questioned whether there are currently measures available to do what is required or whether what is currently being measured is the intended outcome of the interventions/services being delivered.

> *"(I'm) quite conflicted I think because they're really important to show the value of what we're doing, there is no question about that. I think the range of measures in parent-infant mental health are probably not good enough yet." (P07)*

Some respondents suggested that while measures of parental mental health and infant development are often used by services, these are secondary to the relationship, which is perceived to be the primary outcome for parent-infant relationship services.

> *"We're not there to improve mum's mental health, absolutely do we think that a good intervention is likely to have a positive impact, yes. But that's not our aim, so should we even be measuring that?" (P04)*

> *"Though we collect maternal mood measures, these are less relevant to the main work we are doing. I use goal-based ratings, but again this is rarely a measure of the relationship." (Survey)*

> *"ASQ-SE…we've stopped using that because it was useful at measuring the infant's development but not necessarily related to what was going on."* (P06)

In discussing their experience of trying to find an appropriate outcome measure, several practitioners referred to measures which had been used but did not do what the service required and so use was discontinued.

> *"We were using the SDQ but again it's not something that seemed, certainly when I used it, it didn't seem particularly useful or was sensitive to what I was seeing and working with."* (P05)

> *"I guess we had the KIPs but we didn't, it doesn't do what we want it to do so for us that didn't work."* (P04)

> *"A generic measure often doesn't seem to capture the very individual work we do with individual families."* (Survey)

Many responses stressed the need for a simple measure which can easily be completed with families, however this measure is also required to assess the many complex and subtle factors in the parent-infant relationship. The responses below are examples of the somewhat contradictory expectations which are held about outcome measures, in particular the need to capture multiple constructs.

> *"It would have to be something that's very simple, straightforward and just capturing the sort of key things that we know that are important."* (P03)

> *"Capturing the range of impacts under one measure - e.g., impact on parent confidence, attachment, child temperament."* (Survey)

Some practitioners did acknowledge this tension, with one summarising the issue.

> *"Because in a way what we're looking for is this straightforward measure but to measure something that's incredibly complicated. So, I think maybe we're all looking for the Holy Grail."* (P04).

None of the respondents in the survey or interviews suggested they had found a measure they were satisfied with, nor were they aware of other services where this was the case. Comments inferred a need to manage expectations as to what outcome measure can do. Perhaps, as the respondent below suggests, if no team has yet found a solution, then finding one measure is not feasible;

> *"I've not found any team that have said, 'Oh, yes, we're using this and it's brilliant and it works really well, and it measures exactly what we want and it's great.' "* (P05)

It is unlikely that a single measure could be developed to meet all these needs, and certainly not a measure which would be quick and simple to complete. Potentially it would be more helpful to acknowledge this complexity and for practitioners and commissioners to collaborate on realistic ways to evidence change in the domains which matter most to the service.

2 Quantifying the worth of the work

One reason for the importance of measures alluded to by respondents was to provide quantifiable evidence, primarily to funders and commissioners, of the effectiveness of the interventions they delivered and of the value of services – these are often intrinsically linked.

> *"What can be measured is usually more valued and better funded." (Survey)*

In several responses, outcome data was described as a source of evidence to validate the effectiveness of the work being done and thus encourage commissioners to fund the service. This was often described in terms of outcome data being used to communicate the value of a service to commissioners in a language which they understand.

> *"You're just selling it to commissioners, you know, it would help us validate if you've got a tool as to why we're going in." (P03)*

> *"Data is useful for service managers and commissioners to understand the value of the services they are being offered." (Survey)*

> *"In order to evidence the value of a service, this is the language commissioners talk." (Survey)*

The use of outcome data as proof or evidence of a services efficacy was raised by several respondents, it was inferred that this was being driven by commissioners or service managers rather than the practitioners themselves. Although the need to evidence outcomes of services was acknowledged as necessary by practitioners, several comments framed this as a them/us divide in motivating factors.

> *"I think ideally they want outcome measures for commissioning and so on to prove that we're valuable and we are doing effective work." (P05)*

> *"From a commissioning perspective, some kind of tangible, easily condensed outcome data is important for evidencing the efficacy of our service." (Survey)*

In some responses, the perceived drive to evidence the value of a service through quantitative outcome data was viewed a need to gather a paper trail of data.

> *"Our manager came in and she said like, "How are you measuring it with the baby? How are you-- like show me on paper where you've got it with the baby"" (P08)*

The emphasis placed on quantifying the outcomes of a service was described, in some responses, as being at the expense of qualitative, often framed as "soft" measures. Placing a numerical value on parent-infant relationship outcomes was expressed as being at odds with professional expertise in observing the relationship between parent and infant, and it was suggested that the prioritisation of quantitative data could reduce the value of nuanced analysis of the relationship.

> *"soft measures, like our observations, and so our clinical records are really, really important."* (P03)

> *"Outcome measures are tasks completed not the quality of relationships."* (Survey)

Some responses indicated that the devaluing of qualitative data and converse prioritising of quantitative data could lead to difficulty in demonstrating outcomes using qualitative data, and referred to the need to also give appropriate weight to qualitative data which is being collected by services.

> *"It's sad that there's just so much onus on what number they were at the beginning and what number they were at the end rather than everything else that they mean."* (P01)

> *"I feel they (measures) should also never be given too much weight and other 'outcomes' such as feedback, thank you cards, reduction in child protection concerns etc are of equal significance when working with families."* (Survey)

> *"Charity based services constantly face closure because it's so difficult to demonstrate clear objective outcomes despite rich qualitative data."* (Survey)

These responses appear to show a tension for practitioners in recognising that quantitative outcome measures can be useful in their work, primarily driven by the perception that these can demonstrate service value objectively but to the detriment of qualitative data collection.

3 Seeing change - the importance of the visual

The use of visual metaphors in practitioners' language around measures was a striking feature of the responses. This was demonstrated in terms of whether change could be 'seen' when looking at outcome measures across the course of an intervention with a family whilst observing what was happening in the parent-infant relationships and in the therapeutic space.

Many responses referred to a variation on the idea of measures as a tool with which change can be "seen" to show progress across the course of an intervention, as a visual tool for commissioners as one response suggested;

> *"We use our ROMs (Routine Outcome Measures) to illustrate positive change." (Survey)*

Several responses focused on a perceived need for the families to "see" that they are making progress,

> *"(there is) therapeutic benefit for clients to 'witness' their progress in black and white!" (Survey)*

Several practitioners expanded on this, suggesting there was value for the client in being shown the records of their language and how this may have shifted across the course of an intervention.

> *"I did that at the start and then I've done that a little way on through the work and at the end of work and can really see the difference and show the person, 'Look, this is how you were feeling about your baby when you started, and this is how you're feeling now'." (P04)*

> *"it's so nice to do it, just reading it to them and going through it with them because you can say, 'Remember when I was here in March and we had this conversation and remember how far you--, just look at how far you've come'." (P08)*

This suggests that there is a reflective value for the clients and practitioners to "look" at change over the course of an intervention, and that outcome measures can provide relatable evidence of the change to the client, thus being able to visualise this more clearly.

Equally, seeing change in the parent and infant relationship was highly valued by many practitioners. Despite the survey indicating limited use of observational outcome measures, professional skills in observing the parent-infant relationship were central to practitioner's view of what they are aiming to achieve with families.

> *"It would be a change in terms of how I observe the parent/infant relationship. So that I would be seeing a parent that's moved towards being more attuned, more relaxed, in the interactions with their infants." (P04)*

Seeing the responses of infants and their relationship with their parent is crucial in "hearing the voice of the child" in therapeutic work when the "client" is too young to fully verbalise, observations by practitioners may be essential for monitoring progress in therapeutic work.

This is also important in differentiating the perspectives of the parent given verbally, and the behavioural indications of the infant. As one respondent points out, what practitioners need to see is;

> *"how is the baby responding and behaving and can we see changes in how baby is with the parent and in general in themselves through the work." (P05)*

Despite the importance for practitioners of their observations of the parent-infant relationship in the therapeutic space, there was a challenge in recording these observations and using them as a measure of the service. Although respondents described recording their observations in clinical notes, several stated the need for a more formalised mechanism for doing so;

> *"it would be useful just recording some of those observations of parent/child interactions [that] would be very helpful." (P03)*

> *"But I think the thing we struggle with is just not having a kind of measure that we can use that's really about our observation of the parent-infant relationship and where it started and where it's at, at the end of a piece of work." (P07)*

Many comments suggested that observation is central to the work carried out by practitioners, and that more value could be given to clinical observations of the parent-infant relationship recorded in standard clinical practice. However, the reported use of validated observational measures such as the Parent Infant Relational Assessment Tool (PIRAT) and the Strange Situation Procedure (SSP) in the survey was low, and in the qualitative responses this was often attributed to time needed to code observations (see Theme 5). The centrality of observation coupled with the low usage of validated observation measures, indicates a need for recording of clinical observations which can be utilised as outcome measures without creating additional burden for staff or clients.

4 The therapeutic use of measures

The use of measures as a therapeutic or clinical tool, rather than pre/post measures for assessing outcomes, was a frequent feature in the responses. Some measures were described as tools to approach or explore potentially sensitive issues and may aid the development of a treatment plan. For others, they served to aid the practitioner's understanding of the client.

> *"When I complete ROMS with each family we look over and discuss them together to help ensure they are correct, and they can be a great conversation starter for some more difficult questions."* (Survey)

> *"A lot of the measures that we use generate these conversations that are part of our assessment that are really important, so I think it's good to have those even when they're difficult."* (P08)

The distinction of measures, not only to evaluate the outcomes of a programme of work with a family but to make initial assessments and formulate the therapeutic programme, suggested that measures were perceived to simultaneously have diverse roles in service delivery.

> *"So, I would say that the measures help us to formulate a little bit. I think some of the measures rather than being outcome measures are more assessment measures really."* (P07)

> *"Can be helpful to use together with the client as part of the process of assessment/ thinking and treatment."* (Survey)

The use of measures at the start of an intervention as a tool to assess the needs of the parent/infant dyad or to plan the therapeutic intervention, suggests that there needs to be a clearer distinction around the purpose of a given measure, whether it functions as an initial screening or as a measure of impact following intervention.

In addition to opening conversations and developing treatment plans, several responses valued the exploratory use of measures for understanding what is going on for a family during the process as an aid to the clinical work, as one practitioner phrased it;

> *"it can be useful as a measure but it can also be useful as a kind of clinical tool as, to well, to kind of aid that exploration."* (P04)

One example of the diverse application of measures was highlighted in a practitioner's use of an approach to measurement;

> *"So something like KIPS which we used we thought it was a useful tool but not necessarily as a measure. Using video, video work was actually therapeutically useful but not necessarily measuring."* (P06)

These responses suggest that measures are used as a way for seeing where a particular family is and what their needs are, as well as opening further discussions suggesting that measures which are perceived as clinically useful may be more favourable to practitioners.

> *"All the tools we use therapeutically, so even the HADS." (P07)*

However, the multiple use of measures for different functions again shows the high level of expectations around what an outcome measure can do which was discussed in Theme 1, and perhaps this lack of clarity is a barrier to standardisation of outcomes across services.

5 Feeling the burden – time and space to complete measures

Responses often touched on the practitioners concerns about the burden that outcome measures place on both clinicians and parents; several comments referred to the time required to complete and score measures particularly the coding of observational measures.

> *"Time for scoring any lengthy measures is non- existent in a busy team."* (Survey)

> *"it's just time, isn't it, with things like that that just, if you're rating five minute interaction, it can take you three hours to kind of rate something like that."* (P08)

Although practitioners recognise the value of some of the outcome measures, those that were complex observational measures were suggested to be overly time intensive which made their use problematic/impractical.

> *"We are actually being pushed at the moment to try and use the PIRAT. which is amazing. Lovely relational tool. It just takes so long. So long."* (P01)

> *"I am sure it would be really useful, but it (CARE-Index) doesn't seem like it would fit in time-wise to use that routinely with the families that we see."* (P05)

Many respondents viewed outcome measures as an addition to their existing workload,

> *"Staff are bombarded with forms to fill in, reports to write, always another form, always playing catch-up."* (Survey)

While respondents were keen to use measures appropriate to their client's needs, an important factor in the selection of measures was the time requirements for completion, as such those measures (e.g. PIRAT) which were perceived as taking longer were suggested to be within the capacity of the service.

> *"Because the reality is that you need to find something you can do relatively quickly because it's just not feasible to have something that takes, you know, four hours for each family."* (P04)

> *"having a fairly brief outcome measure that uses positive, so strength-based or language that can be meaningful to families."* (Survey)

Similarly, several survey responses around the barriers to the use of outcome measures indicated a perception of outcome data collection as a potentially negative act towards families with emotive language suggesting "bombarding" or "persecuting" families.

> *"Other services families are engaged with also use measures and families can feel bombarded."* (Survey)

> *"Timing of asking (is a challenge) as families often want to tell their story or are highly stressed/feel persecuted by being asked to fill questionnaires in."* (Survey)

The clients who are referred to parent-infant relationship services may be vulnerable for a variety of reasons; several respondents acknowledge their discomfort at attempting to complete measures in the early stages of the therapeutic work due to these vulnerabilities.

> *"(there is the) difficulty of finding time to input data and the space to go through questionnaires when parents who are newly referred are often in desperate states." (Survey)*

Some responses also raised concerns about the appropriateness of completing measures in a therapeutic context and the duality of the therapist/evaluator role that this can lead to, with some respondents reporting that completing measures created a disruption to their therapeutic role.

> *"it can feel a bit like needing to be somebody else in order to ask some of those questions." (P07)*

> *"it's really hard to stop the flow, the therapeutic flow and say, 'Sorry, but we have to do these measures.' " (P06)*

Practitioners also suggested measures need to be carefully timed so as not to threaten the therapeutic relationship between therapist and client, and one suggested that outcome data collection can feel like a potentially damaging act;

> *"to sort of go in armed with a lot of questionnaires before you've built a relationship feels quite difficult really." (P07)*

In the responses there appeared to be a perceived burden on the practitioner to manage the perceived negative impacts on parents in completing the measures, whether that is managing potentially stigmatising language, managing potential threat to the relationship, or waiting to complete measures when a client is in a less distressed state. Overall, within this theme there was a somewhat negative perception of measures in practice, with some respondents describing measures as intrusive both emotionally and temporally for practitioners and parents. However, it would be useful to know whether this perception is echoed in clients' experiences of outcome measures, or whether this is mainly attributable to therapists' expectations and their confidence in carrying out measures. Respondents suggested that, while they appreciate the value of observational measures which require video and coding, in practice these are not often feasible due to time constraints.

6 Towards shared understanding and practice

Despite the suggested tensions between stakeholders' perception of purpose and use of measures, the practitioners' responses suggested a drive to develop a shared understanding and practice.

Many responses indicated the need for measures to be meaningful, not just to evidence outcomes for commissioners but to support the work done by practitioners;

> *"I feel that the CCG (Clinical Commissioning Group) and commissioners put pressure on practitioners to complete ROMS in order to simply collect data and this can mean that their true meaning (to help the families and practitioners) can get lost."* (P01)

> *"Lack of insight from leadership teams who make decisions about which tools to use means the impact of the work isn't captured."* (Survey)

These examples suggest a collaborative approach to selecting the most appropriate measures for a service could be taken, with discussion between commissioners and providers, to develop mutually meaningful measures of the intended outcome.

> *"Having a good understanding of what the outcome measure is asking and why it is being completed helps facilitate the use of the measures on a service level."* (Survey)

The desire/drive of practitioners to share practice and to develop a shared set of measures, both within teams and across services delivering parent-infant relationship and infant mental health work, was shown in several responses to questions around good practice;

> *"In order to try and embed them in our service and try and sort of get some reliability and consensus around within the team about how we're using them both we've been meeting as a team quite regularly"* (P08)

> *"We, as a team with our particular training and background and thinking about the parent/infant relationship, felt differently about what was important and what was acceptable, and what was positive in an interaction, as opposed to what KIPs was telling us."* (P07)

However, several responses indicated there is not yet consensus amongst services, or indeed practitioners, on a clear definition of key concepts contained in the terms "Parent-infant relationships" and "infant mental health" or how they intersect with other aspects of child development. As this respondent explains, if there is not consensus on something, how can it be measured;

> *"Lack of general recognition in society, and even amongst professionals, of the meaning of the term 'infant mental health'. It is hard to assess a factor the constituents of which are not generally agreed upon."* (Survey)

> *"Demonstrating the link between improved parent-infant relationships and child development. I still feel people find it hard to comprehend that improved attachment impacts on development."* (Survey)

As well as the broader need for understanding over those concepts which may be included in outcome measures for parent-infant relationship services, i.e. "infant mental health", practitioners suggested there is also a need for;

> *"agreeing on a common approach to outcome measurement across the service or services."* (Survey)

The value of a shared approach was suggested to be across services;

> *"We have tried several different outcome measures and would like to be consistent and use an approach that is more universally acceptable."* (Survey)

> *"I'd love it if all parent-infant work, we all could work together on measures that were suitable across countrywide."* (P06)

> *"learning from other areas where they've implemented these things and creating a little bit of momentum would be really good because it's just isolated pockets at the moment and it's almost like that needs to come together to create some sort of sense of change."* (P03)

Moreover, a shared understanding between therapist and client in order to measure the work in a way which is meaningful to both parties. Recommendations included working collaboratively, and tailoring measures, in this respect Goal Based Outcomes were praised by some respondents;

> *"I am not always sure how much value families take from the scores and if it adds to their growing understanding."* (Survey)

> *"When it's your goal and you've created it together, it is more personal. It means more to them. It means more to me."* (P01)

> *"So I want to use an outcome measure that's going to be meaningful to them and that we can use collaboratively and that is going to be helpful to them."* (P05)

One suggestion for how to increase the likelihood of measures which are meaningful to all parties is to select, or even develop, outcome measures with parents in mind, or in collaboration with parents.

> *"This doesn't seem to have really been made with parents in mind, more clinicians in mind and measuring people in mind but less in the trying to help people together with people."* (P05)

> *"What it does is helps practitioners to think about agreeing on intended outcomes with the families that they support and using those to measure progress."* (P02)

However, ensuring that measures are 'meaningful' will contribute to the challenge of selecting measures, as what is 'meaningful' is subjective and potentially different for all stakeholders. Nonetheless, the need for the parents' perspective on the meaning of outcome measures for them is a factor which may have been given little attention in the implementation of evaluation and outcome measurement, and for future good practice may need to be given more weight.

7 Thinking long term

One aspect of outcome measurement which several respondents felt was underexplored, was the opportunity to measure over the long term to be able to infer sustained change. Some practitioners' comments suggested this provided value to their work, and had built this into their practice;

> *"The plan…is to go back and do a follow-up because obviously what we want to see as well is, did this intervention last, you know? Was there a kind of sustained change over time in terms of the quality of that relationship and those outcomes."* (P04)

> *"What would be really helpful I think is to be able to follow families up in a couple of years and then do some sort of measure, questionnaire with them to find out, to try and track whether what we offer as well is effective longer term."* (P05)

Taking a more long-term approach to collecting outcome data was also mentioned to support the sustainability of service delivery and knowledge about whether an intervention is creating the desired impact. Some respondents discussed the pitfalls of taking a short-term approach as they felt this did not give services a realistic opportunity to be assessed;

> *"Pressure to provide outcomes data before progress is always observable"* (Survey)

> *"The Family Nurse Partnership was decommissioned almost at the time when it was going to find out whether it was making a difference."* (P06)

A short-term approach was also perceived as a problem for implementing suitable measures, with several respondents reporting that there is high turnover in mandated measures which may limit the chances of finding whether both the measure and the service is effective.

> *"Well we'll get trained in this and next year it'll be something different. It's just bonkers."* (P01)

> *"shall we implement and invest in that, and then it just sort of goes out the window."* (P03)

The perceived lack of a long-term commitment to measures or services, and responses which detailed frequent changes of measures being implemented, may challenge the ability of services to invest their time and energy in the measure of the moment. The need for consolidation rather than transformation was evident in some responses, as this practitioner stated;

> *"I think at the moment we need to just get on with embedding what we've got."* (P07)

In the current socio-political landscape where practitioners and commissioners are working in the context of short-term funding cycles and political uncertainty it may be difficult to advocate for long term investment in outcome measures, however if the important work which is being done is to be valued this is a crucial goal.

Section 3:

Conclusions from the study - evidence and practice

The qualitative responses highlighted the array of tensions which exist around outcome measures, manifest in conflicting agendas, the need to satisfy a range of sometimes contradictory needs, and the lack of consensus on what to measure and how to measure it. This echoes the learning from the literature which indicated there is not consensus on either the constructs to be measured (i.e. parent-focused confidence, mental health, form of parent-child relationship etc.), or the outcome measures with which to do this.

Practitioner's responses indicated that outcome measures are primarily viewed as evidence or proof that change has happened, however there are a range of stakeholders who need to "see" change and how it can best be evidenced. This suggests that there needs to be discussion at the earliest possible stage between stakeholders about the outcome which the service is aiming to change, and from this to select the most feasible measure of this. This discussion could be focused around the development of a robust theory of change for services as recommended by the Early Intervention Foundation. [4]

The difficulty in measuring objective change is, in part, due to the highly individual nature of the work being done which varies not just between services, but is tailored for each family engaged with. The qualitative responses indicated that one area of change agreed by most practitioners is an improvement in the relationship between parent and infant, and that other measures are secondary to this. While the improvement in relationship is often framed as an improvement in attachment, measuring this may not be feasible given the constraints of service delivery.

While observational measures are often seen as the "gold standard" for measuring parent-infant relationship (Gridley et al., 2019), the researchers gained a sense that so called "gold standards" can be self-perpetuating, as in systematic reviews where studies are often only included if the methodology follows a RCT design, often including observational measures. Trials adhere to this standard and then are reported on in systematic reviews perpetuating a narrow set of norms. Thus, the academic evidence base tends to prioritise interventions which have been trialled using observational measures and so this type of measure can be perceived as "best practice" through frequent practice. While Bagner et al. (2013) state that "A more comprehensive multi-method evaluation would ideally include both questionnaires and behavioural observations" this is difficult to achieve in clinical practice. While a "comprehensive evaluation" may be necessary in clinical trials or research studies into the efficacy of specific interventions, it could be argued this is excessive for services doing wide ranging clinical work with parents.

[4] Step 1 Creating a theory of change - EIF Evaluation Hub

It was also suggested in the practitioner responses that while services are measuring other constructs, such as parental mental health and that this is inextricably linked with infant mental health and the parent-infant relationship, these measures should not be reported as a proxy for change in the parent and child relationship.

The study also highlighted the limited number of measures suitable for younger children, particularly those under 12 months. While many parent-infant relationship services provide support for ages 0-5 and in some cases link with CAMHS, many parent report measures are dependent on the infant having reached certain developmental milestones, i.e. the ability to smile, and so are not suitable for the very young. This is a significant gap in the policy and practice of evaluating parent-infant relationship interventions as many services aim to address disruption to relationship in the first 1001 days.

The high level of demands placed on what outcome measures can achieve was a common thread across the study, suggesting that the development of a shared understanding about the purpose of measures may support more realistic expectations for all stakeholders. If services can identify a primary outcome and focus on the measurement of this rather than attempting to implement a large range of measures to cover all possible outcomes, this would be beneficial.

As well as the high level of demands on what outcome measures can achieve, the responses to both survey and interviews showed that there is a high burden already felt by those delivering parent-infant therapeutic interventions due to limited time and staffing resources. The time needed to complete more complex, observational, measures was not felt to be available to practitioners, and concerns were also raised about the perceived burden which may be felt by families who are asked to complete many measures in the form of questionnaires at a time when they are in distress.

Further work with stakeholders, especially including the voice of parents and practitioners, to develop the use of outcome measures may help to address some of the negative assumptions which were reported. The responses from practitioners showed that they value the work they do, and they would value a robust way of evidencing this, drawing on their practical expertise and the evidence from research.

Recommendations for practice

Five recommendations for practice have been drawn from the learning in the report which can be used to guide the development of outcome measures for parent-infant relationship services.

1 Being Realistic

There is no single measure, or even a set of measures, which can be all things to all people. It became clear during this study there is an ardent desire for a simple, easy to use and universally recommended measure, and that through it we might identify just such a measure. However, we returned multiple times to the conclusion, it is not possible to measure the vast range of potential parent-infant relationship outcomes and, in attempting to do so, we may be doing a disservice to the complexity of the work to support parent-infant relationships.

2 Seeking Clarity

There is a lack of clarity around what the primary outcome should be, and the suitability of measures for outcome measurement as opposed to screening/ assessment. If it can be agreed that the primary outcome for services is an improvement in the parent-infant relationship, then we need clarity as to what aspects of that can be captured in a measure and should be cautious of the use of proxy measures.

3 Capturing Observation

Responses from practitioners showed a real focus on "seeing" the child and the relationship, and how their expertise in this is crucial to the work done with families but is often not captured as part of evaluation. Given the "gold standard" focus on observational measures of attachment, there is value in committing resources to implement and carry out observational measures to see what happens in the parent-infant relationship during, and following, intervention. It should be noted that some interventions have this built-in in an informal manner, for example, Video Interaction Guidance (VIG).

4 Thinking Long-term

To understand the impact on child outcomes, given the complex funding landscape and drivers to deliver evidence in a short period of time, it is necessary to challenge short-term thinking. This extends across giving time for services to do the work and evaluations to capture the impact, but also in implementing measures and then using them consistently.

5 Working Together

A range of stakeholders are involved in the development and delivery of parent-infant relationship work: practitioners, parents, researchers, commissioners, service managers, and national bodies, and it is important these voices are all heard and valued in the identification of measures and development of evaluation. Accepting there will be competing priorities, and that no measure can be all things to all people, is a good starting point from which to build a shared understanding, and to underpin the implementation of the other four learning points of this review.

References

References Measures

1. Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). Harcourt.
2. Biringen, Z. (2008). The Emotional Availability (EA) Scales and the Emotional Attachment & Emotional Availability (EA2) Clinical Screener (4th edition): Infancy/Early Childhood Version
3. Briggs-Gowan, M. J., Carter, A. S., Irwin, J. R., Wachtel, K., & Cicchetti, D. V. (2004). The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology,* 29, 143–155.
4. Brockington, I. F., Fraser, C., & Wilson, D. (2006). The Postpartum Bonding Questionnaire: a validation. *Archives Of Women's Mental Health*, 9(5), 233–242.
5. Cox, J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British Journal Of Psychiatry: The Journal Of Mental Science*, 150, 782–786.
6. Crittenden, P. (2004) *CARE-Index: Manual*. Family Relations Institute.
7. Feldman R. (1998). Coding Interactive Behaviour manual. Israel: Bar-llan University: Unpublished manuscript
8. Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal Of General Internal Medicine,* 16(9), 606–613.
9. Law. D., (2011) *Goal Based Outcomes (GBOs): Some useful information*. Internal CORC publication, Available at http://www.corc.uk.net (Retrieved August 2021)
10. NICHD Early Child Care Research Network. (2006) Infant–mother attachment classification: risk and protection in relation to changing maternal caregiving quality. *Developmental Psychology*, 41: 38–58.
11. Oates, J., Gervai, J., Danis, I., Lakatos, K., & Davies, J. (2018). Validation of the Mothers' Object Relations Scales Short-form (MORS-SF). *Journal of Prenatal and Perinatal Psychology and Health*, 33(1), 38–50
12. Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, 166(10), 1092–1097.
13. Squires, J., & Bricker, D. (2009). *Ages & Stages Questionnaires, Third Edition* (ASQ-3). Baltimore, MD: Brookes Publishing
14. Squires, J. K., Waddell, M. L., Clifford, J. R., Funk, K., Hoselton, R. M., & Chen, C.-I. (2013). A Psychometric Study of the Infant and Toddler Intervals of the Social Emotional Assessment Measure. *Topics in Early Childhood Special Education*, 33(2), 78–90.
15. Stocky, A. J., Tonge, B. J., & Nunn, R. J. (1996). The reliability and validity of the Bethlem Mother-Infant Interaction Scale. *The British Journal of Psychiatry: the Journal of Mental Science*, 169(1), 27–29.
16. Svanberg P. O., Barlow. J., & Tigbe. W. (2013) The Parent–Infant Interaction Observation Scale: reliability and validity of a screening tool. *Journal of Reproductive and Infant Psychology*, 31: 1–10.
17. Zeanah, C.H., Carter, A., Cohen, J., Egger, H., Keren, M., Gleason, M.M. et al. (2015). DC:0–3 to DC:0–3R to DC:0–5: A new edition. *ZERO TO THREE*, 35, 63–66.

References General

1. Adams, L., S. Morris, S., Svanaes, K., Beninger, K., Smith, A., Gellen, S & Jones, H (2021). *Trial evaluation protocol: Evaluation of empowering parents and professionals using Video Interactive Guidance (VIG) programme* (Babies 1st). Technical Report. What Works Centre for Children's Social Care
2. Ali, E., Letourneau, N., & Benzies, K.M. (2021). Parent-Child Attachment: A Principle-Based Concept Analysis. *SAGE Open Nursing*, 7.
3. Armstrong, V. G., Dalinkeviciute, E., & Ross, J. (2019). A Dyadic Art Psychotherapy Group for Parents and Infants: Piloting Quantitative Methodologies for Evaluation. *International Journal of Art Therapy*, 24(3), 113-124.
4. Bagner, D. M., Rodríguez, G. M., Blake, C. A., Linares, D., & Carter, A. S. (2012). Assessment of behavioral and emotional problems in infancy: A systematic review. *Clinical Child and Family Psychology Review*, 15, 113–128.

5. Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological bulletin*, 129(2), 195–215.
6. Barlow, J., Bennett, C., Midgley, N., Larkin, S. K., & Wei, Y. (2015). Parent infant Psychotherapy for Improving Parental and Infant Mental Health: A Systematic Review. *Campbell Systematic Review,* 11(1), 1–223.
7. Barlow, J. (2012). Child maltreatment during infancy: atypical parent–infant relationships. *Paediatrics and Child Health*, 22(11), 465–469.
8. Barlow, J., Sleed, M., & Midgley, N. (2021). Enhancing parental reflective functioning through early dyadic interventions: A systematic review and meta-analysis. *Infant Mental Health Journal*, 42(1), 21–34.
9. Booth, A.T., Macdonald, J.A., & Youssef, G.J. (2018). Contextual stress and maternal sensitivity: A meta-analytic review of stress associations with the Maternal Behavior Q-Sort in observational studies. *Developmental Review*. 48 145-177
10. Cohen, N.J. Lojkasek, M., Muir, E., Muir, R., & Parker, C.J. (2002). Six-month follow-up of two mother-infant psychotherapies: Convergence of therapeutic outcomes. *Infant Mental Health Journal*. 23, 361-380.
11. Condon, J. (2012) Assessing attachment, a work in progress: to look, to listen or both, *Journal of Reproductive and Infant Psychology*, 30:1, 1-4,
12. Crncec, R., Barnett, B., & Matthey, S. (2010). Review of scales of parenting confidence. *Journal of nursing measurement*, 18(3), 210–240.
13. Domoney, J., Fulton, E., Stanley, N., McIntyre, A., Heslin, M., Byford, S., Bick, D., Ramchandani, P., MacMillan, H., Howard, L. M., & Trevillion, K. (2019). For Baby's Sake: Intervention development and evaluation design of a whole-family perinatal intervention to break the cycle of domestic abuse. *Journal of Family Violence*, 34(6), 539–551
14. Fagan, M. A., Frosch, C. A., Middlemiss, W., LaCoursiere, J. N., Owen, M. T., Hane, A. A., & Welch, M. G. (2019). The practical utility of the Welch Emotional Connection Screen for rating parent–infant relational health. *Infancy*, 24(6), 881–892.
15. Fonagy, P., Sleed, M., & Baradon, T. (2016). Randomized controlled trial of parent-infant psychotherapy for parents with mental health problems and young infants. *Infant Mental Health Journal*, 37(2), 97–114.
16. Gridley, N., Blower, S., Dunn, A., Bywater, T., Whittaker, K., & Bryant, M. (2019). Psychometric Properties of Parent–Child (0–5 years) Interaction Outcome Measures as Used in Randomized Controlled Trials of Parent Programs: A Systematic Review. *Clinical Child and Family Psychology Review*, 22(2), 253–271.
17. Hall, C. L., Moldavsky, M., Baldwin, L., Marriott, M., Newell, K., Taylor, J., Sayal, K., & Hollis, C. (2013). The use of routine outcome measures in two child and adolescent mental health services: a completed audit cycle. *BMC psychiatry,* 13, 270.
18. Henderson, M., Wittkowski, A., McIntosh, E. et al. (2019). Trial of healthy relationship initiatives for the very early years (THRIVE), evaluating Enhanced Triple P for Baby and Mellow Bumps for those with additional social and care needs during pregnancy and their infants who are at higher risk of maltreatment: study protocol for a randomised controlled trial. *Trials* 20, 499
19. Hendriks, A., Van der Giessen, D., Stams, G., & Overbeek, G. (2018). The association between parent-reported and observed parenting: A multi-level meta-analysis. *Psychological Assessment*, 30(5), 621–633.
20. Høifødt, R. S., Nordahl, D., Pfuhl, G., Landsem, I. P., Thimm, J. C., Ilstad, L. K. K., & Wang, C. E. A. (2017). Protocol for the Northern babies longitudinal study: predicting postpartum depression and improving parent–infant interaction with The Newborn Behavioral Observation. *BMJ Open*, 7(9), e016005–e016005
21. Holt, C., Gentilleau, C., Gemmill, A. W., & Milgrom, J. (2021). Improving the mother-infant relationship following postnatal depression: a randomised controlled trial of a brief intervention (HUGS). *Archives of women's mental health,* 24(6), 913–923.
22. Hutchings, J., Griffith, N., Bywater, T., & Williams, M. E. (2017). Evaluating the Incredible Years Toddler Parenting Programme with parents of toddlers in disadvantaged (Flying Start) areas of Wales. *Child: care, health and development,* 43(1), 104–113.
23. Kersten Alvarez, L. E., Hosman, C. M., Riksen-Walraven, J. M., Van Doesum, K. T., & Hoefnagels, C. (2011). Which preventive interventions effectively enhance depressed mothers' sensitivity? A meta-analysis. *Infant Mental Health Journal*, 32(3), 362-376.
24. Kohlhoff, J., Cibralic, S., Wallace, N., Morgan, S., McMahon, C., Hawkins, E., Eapen, V., Briggs, N., Huber, A., & McNeil, C. (2020). A randomized controlled trial comparing parent child interaction therapy - toddler, circle of security- parenting™ and waitlist controls in the treatment of disruptive behaviors for children aged 14-24 months: study protocol. *BMC Psychology*, 8(1), 93.

25. Kristensen, I. H., Juul, S., & Kronborg, H. (2020). What are the effects of supporting early parenting by newborn behavioral observations (NBO)? A cluster randomised trial. *BMC psychology,* 8(1), 107.
26. Lotzin, A., Lu, X., Kriston, L., Schiborr, J., Musal, T., Romer, G., & Ramsauer, B. (2015). Observational Tools for Measuring Parent–Infant Interaction: A Systematic Review. *Clinical Child and Family Psychology Review*, 18(2), 99–132.
27. Lowell, D.I., Carter, A.S., Godoy, L., Paulicin, B., Briggs-Gowan, M.J. (2011). A Randomized Controlled Trial of Child First: A Comprehensive, Home-Based Intervention Translating Research into Early Childhood Practice. *Child Development,* 82(1), 193-208.
28. Maas, A. J. B., Vreeswijk, C. M. J., de Cock, E. S., Rijk, C. H. A., & van Bakel, H. J. (2012). “Expectant Parents”: Study protocol of a longitudinal study concerning prenatal (risk) factors and postnatal infant development, parenting, and parent-infant relationships. *BMC Pregnancy and Childbirth*, 12(1), 46–46.
29. Malinit, J. (2021). The outcome of an attachment-based infant mental health therapeutic play programme on infant temperament, parent-infant relationship & maternal reflective functioning. *BJPsych Open*, 7(S1), S268–S268.
30. Mascheroni, E., & Ionio, C. (2019). The efficacy of interventions aimed at improving post-partum bonding: A review of interventions addressing parent-infant bonding in healthy and at-risk populations. *Journal of Neonatal Nursing* 25(2), 61–68.
31. McLuckie, A., Landers, A. L., Curran, J. A., Cann, R., Carrese, D. H., Nolan, A., Corrigan, K., & Carrey, N. J. (2019). A scoping review of mental health prevention and intervention initiatives for infants and preschoolers at risk for socio-emotional difficulties. *Systematic Reviews*, 8(1), 183–183.
32. Menashe-Grinberg, A., Shneor, S., Meiri, G., & Atzaba-Poria, N. (2021). Improving the parent-child relationship and child adjustment through parental reflective functioning group intervention. *Attachment & Human Development,* 1–21.
33. Mihelic, M., Morawska, A., & Filus, A. (2017). Effects of early parenting interventions on parents and infants: A meta-analytic review. *Journal of Child and Family Studies*, 26(6), 1507–1526.
34. Mirza, S. (2020) *Exploring the use of Infant CARE-Index to assess parent-infant interactions (0-15 months): Measurement properties and acceptability in practice*. MSc by research thesis, University of York.
35. Mothander, R., Furmark, C., & Neander, K. (2018). Adding “Circle of Security – Parenting” to treatment as usual in three Swedish infant mental health clinics. Effects on parents’ internal representations and quality of parent-infant interaction. *Scandinavian Journal of Psychology,* 59(3), 262–272.
36. Murray, L., Fiori-Cowley, A., Hooper, R., & Cooper, P. (1996). The impact of postnatal depression and associated adversity on early mother-infant interactions and later infant outcome. *Child development,* 67(5), 2512–2526.
37. Naughton, A., Perkins, L., McMinn, B., & Kemp, A. (2019). Using an observation tool (Parent-Infant Interaction Observation Scale) to assess parent-infant interaction in the first 2 weeks of life: A feasibility study. *Child: care, health and development*, 45(2), 271–285.
38. Newton, K., Taylor Buck, E., Weich, S., & Uttley, L. (2022). A review and analysis of the components of potentially effective perinatal mental health interventions for infant development and mother-infant relationship outcomes. *Development and psychopathology,* 34(1), 37–54.
39. NICE (2019) NICE guideline [NG26] Children’s Attachment: attachment in children and young people who are adopted from care, in care or at high risk of going into care.
40. NICE (2016) NICE Quality standard [QS133] Children’s Attachment. Overview | Children’s attachment | Quality standards NICE (Retrieved August 2021)
41. NICE (2020) NICE Clinical guideline [CG192] Antenatal and Postnatal Mental Health. Overview | Antenatal and postnatal mental health: clinical management and service guidance | Guidance NICE (Retrieved August 2021)
42. NICE (2021) Evidence Review O. Emotional attachment. NG194 Evidence review O (nice.org.uk) (Retrieved August 2021)
43. Nicolson, S., Carron, S. P., & Paul, C. (2022). Supporting early infant relationships and reducing maternal distress with the Newborn Behavioral Observations: A randomized controlled effectiveness trial. *Infant mental health journal,* 43(3), 455–473.
44. O’Hara, L., Smith, E., Barlow, J., Livingstone, N., Herath, N., Wei, Y., Spreckelsen, T. F., & Macdonald, G. (2019). Video feedback for parental sensitivity and attachment security in children under five years (Review). *Cochrane database of systematic reviews (Online)*

45. Patel, J., Smith, R., O'Farrelly, C., Iles, J., Rosan, C., Ryan, R., & Ramchandani, P. (2021). Assessing behavior in children aged 12-24 months using the Strengths and Difficulties Questionnaire. *Infancy: the official journal of the International Society on Infant Studies,* 26(5), 724–734.
46. Perrelli, J. G., Zambaldi, C. F., Cantilino, A., & Sougey, E. B. (2014). Instrumentos de avaliação do vínculo entre mãe e bebê [Mother-child bonding assessment tools]. *Revista paulista de pediatria : orgao oficial da Sociedade de Pediatria de Sao Paulo*, 32(3), 257–265.
47. Poobalan, A., Aucott, L., Ross, L., Smith, W., Helms, P., & Williams, J. (2007). Effects of treating postnatal depression on mother-infant interaction and child development: Systematic review. *British Journal of Psychiatry,* 191(5), 378-386.
48. Pontoppidan, M., Nissa, N. K., Pejtersena, J. H., Julianc, M. M., & Væverd, M. S. (2017). Parent report measures of infant and toddler social-emotional development: A systematic review. *Family Practice*, 34(2), 127–137.
49. Popp, L., Fuths, S., & Schneider, S. (2019). The relevance of infant outcome measures: A pilot-RCT comparing baby Triple P positive parenting program with care as usual. *Frontiers in Psychology*, 10, 2425–2425.
50. Priddis, L. E., & Kane, R. (2013). The development of a scale for tuned-in parenting. *Australian Journal of Psychology,* 65(4), 198–205.
51. Puckering, C., McIntosh, E., Hickey, A., & Longford, J. (2010). Mellow babies: A group intervention for infants and mothers experiencing postnatal depression. *Counselling Psychology Review*, 25(1), 28–40.
52. Rance, S. (2005). Attending to early relationship difficulties: Applying the "Watch Wait and Wonder" approach in a pilot project for babies and parents. *Infant Observation*, 8(2), 125–138
53. Raouna, A., Malcolm, R., Ibrahim, R., & MacBeth, A. (2021). Promoting sensitive parenting in "at-risk" mothers and fathers: A UK outcome study of Mellow Babies, a group-based early intervention program for parents and their babies. *PloS One*, 16(2), e0245226–e0245226.
54. Rayce, S. B., Rasmussen, I. S., Væver, M. S., & Pontoppidan, M. (2020). Effects of parenting interventions for mothers with depressive symptoms and an infant: Systematic review and meta-analysis. *BJPsych Open*, 6(1), E9.
55. Skovgaard, A. M., Olsen, E. M., Houmann, T., Christiansen, E., Samberg, V., Lichtenberg, A., & Jørgensen, T. (2005). The Copenhagen County child cohort: design of a longitudinal study of child mental health. *Scandinavian journal of public health,* 33(3), 197–202.
56. Skovgaard A. M. (2010). Mental health problems and psychopathology in infancy and early childhood. An epidemiological study. *Danish medical bulletin*, 57(10), B4193.
57. Slade, A., Holland, M. L., Ordway, M. R., Carlson, E. A., Jeon, S., Close, N., Mayes, L. C., & Sadler, L. S. (2019). Minding the Baby®: Enhancing parental reflective functioning and infant attachment in an attachment-based, interdisciplinary home visiting program. *Development and Psychopathology,* 32(1), 123–137.
58. Sleed, M., Baradon, T., & Fonagy, P. (2013). New Beginnings for mothers and babies in prison: A cluster randomized controlled trial. *Attachment & Human Development*, 15(4), 349–367.
59. Sleed, M., James, J., Baradon, T., Newbery, J., & Fonagy, P. (2013). A psychotherapeutic baby clinic in a hostel for homeless families: Practice and evaluation. *Psychology and Psychotherapy*, 86(1), 1–18.
60. Sprengeler, M. K., Mattheß, J., Eckert, M., Richter, K., Koch, G., Reinhold, T., Vienhues, P., Berghöfer, A., Fricke, J., Roll, S., Keil, T., Ludwig-Körner, C., Kuchinke, L., von Klitzing, K., & Schlensog-Schuster, F. (2021). Efficacy of parent-infant psychotherapy compared to care as usual in children with regulatory disorders in clinical and outpatient settings: study protocol of a randomised controlled trial as part of the SKKIPPI project. *BMC Psychiatry*, 21(1), 118–118.
61. Szaniecki, E., & Barnes, J. (2016). Measurement Issues: Measures of infant mental health. *Child and Adolescent Mental Health*, 21(1), 64–74.
62. Trombetta, T., Giordano, M., Santoniccolo, F., Vismara, L., Della Vedova, A. M., & Rollè, L. (2021). Pre-natal Attachment and Parent-To-Infant Attachment: A Systematic Review. *Frontiers in Psychology*, 12,
63. Tryphonopoulos, P. D., Letourneau, N., & Ditommaso, E. (2014). Attachment and caregiver-infant interaction: a review of observational-assessment tools. *Infant Mental Health Journal,* 35(6), 642–656.
64. Salomonsson, B., Kornaros, K., Sandell, R., Nissen, E., & Lilliengren, P. (2021). Short-term psychodynamic infant-parent interventions at Child health centers: Outcomes on parental depression and infant social-emotional functioning. *Infant Mental Health Journal,* 42(1), 109–123.

65. Væver, M. S., Smith-Nielsen, J., & Lange, T. (2016). Copenhagen infant mental health project: study protocol for a randomized controlled trial comparing circle of security -parenting and care as usual as interventions targeting infant mental health risks. *BMC Psychology*, 4(1), 57.
66. Van Doesum, K. T., Riksen-Walraven, J. M., Hosman, C. M., & Hoefnagels, C. (2008). A randomized controlled trial of a home-visiting intervention aimed at preventing relationship problems in depressed mothers and their infants. *Child development*, 79(3), 547–561
67. Van Ijzendoorn, M. H., Schuengel, C., & Bakermans-Kranenburg, M. J. (1999). Disorganized attachment in early childhood: Meta-analysis of precursors, concomitants, and sequelae. *Developmental Psychopathology*, 11(2), 225–249.
68. Klein Velderman, M., Bakermans-Kranenburg, M. J., Juffer, F., & van IJzendoorn, M. H. (2006). Effects of attachment-based interventions on maternal sensitivity and infant attachment: differential susceptibility of highly reactive infants. *Journal of family psychology* 20(2), 266–274.
69. Wittkowski, A., Garrett, C., Calam, R., & Weisberg, D. (2017). Self-Report Measures of Parental Self-Efficacy: A Systematic Review of the Current Literature. *Journal of child and family studies*, 26(11), 2960–2978.

Appendix 1. Literature Search Strategy

The review aimed to address the following research questions:

1. Which outcome measures are currently used in evaluation of interventions and services which specifically aim to address disruption in the parent-infant relationship or infant mental health in carer-infant dyads?

2. Which outcome measures are recommended for use in parent-infant relationship/ infant mental health interventions and services in current UK policy documents?

Seven online publication databases (CINAHL, PubMed, Medline, PsychInfo, BMJ Online Journals, ProQuest, EBM Cochrane Reviews) were searched for peer reviewed studies and reviews of parent-infant interventions published from 2010-present (this date range was selected in line with our interest in recent practice in the use of outcome measures) using combinations of the following terms (*indicates truncation);

(parent* OR mother OR maternal OR caregiver) (infant OR baby)

(interaction OR relation* OR respons* OR sensitive* OR attach* OR bond*)

(infant mental health)

(Meta-analys*OR review OR intervention)

Individual searches were also carried out for specific programmes or approaches which have been identified for use by the Parent Infant Foundation, or by specialised parent-infant relationship services in the UK, to support parent-infant relationships including;

“Watch Wait Wonder,”

“Incredible Years,”

“parent-infant psychotherapy,”

“video interaction guidance,”

“video interaction feedback”

“Circle of Security”

Manual inclusion/exclusion was carried out through review of abstracts. The following exclusion criteria were applied:

· Interventions which do not explicitly aim to address an aspect of parent-infant relationship or infant mental health
· Studies of parent-child dyads with child over 5
· Studies conducted in NICU or PICU due to the unique needs of this group of parents.

The citation lists of included papers were searched for potentially relevant literature, including non-peer reviewed publications.

Appendix 2. Measures used by practitioners

2.1 Outcome measures which specialised parent-infant relationship services have used

A search for impact/outcome evaluations published by specialised parent-infant relationship services in the UK up to August 2021 identified publications from 7 services, most published in the last 2-3 years thus giving a helpful picture of current use. The measures reported are those used for direct clinical work with parents, although the evaluations also included details of consultations and training for workforce.

Service (year of publication)	**Measures reported**
ABC PIP, Northern Ireland (2019)	Parent Baby Outcome Star Parenting Stress Index (PSI) Hospital Anxiety and Depression Scale (HADS) Parent Feedback forms
DORPiP, Dorset (2021)	Levels of Adaptive Functioning (LOAF) Caregiver Dimension Parent Feedback forms
Leeds infant mental health service (2020)	PHQ9 GAD-7 MORS Session Rating Scale and Outcome Rating Scale Maternal/Paternal Antenatal Attachment Scale (MAAS and PAAS)
Little Minds Matter Bradford (2019-20)	MORS Goal Based Outcome Measure Parent Feedback forms
LIVPIP now PSS (Liverpool) (2017)	Parent Infant Relationship Global Assessment Scale (PIRGAS) Keys to Interactive Parenting Scale (KIPS) Hospital Anxiety and Depression Scale (HADS) Parent Feedback forms
NEWPIP now Little Minds in Mind – Newcastle, (2021)	Qualitative Interviews
Together with Baby - Essex (2021)	MORS Hospital Anxiety and Depression Scale (HADS) but changed to CORE-10 Parent Feedback forms

2.2 Measures used by practitioners (survey respondents)

Ainsworth Strange Situation Procedure (SSP)

Ages and Stages Questionnaire (ASQ)

Ages and Stages Questionnaire – Social and Emotional (ASQ-SE)

Bayley Scales of Infant and Toddler Development (BSID)

Behaviour Rating Inventory of Executive Function, Preschool Version (Brief P)

Brief Parental Self Efficacy Scale

Child Language Measures

Childhood Trauma Questionnaire, CTQ

Client Service Receipt Inventory (CSRI)

Clinical Outcomes in Routine Evaluation – 10 item (CORE-10)

CORE 6D – complex scoring of six items of the Clinical Outcomes in Routine Evaluation Outcome Measure

Denver II Developmental Screening Test

Difficulties in Emotion Regulation Scale, DERS

Early Attachment Observation (EAO)

Early Executive Functions Questionnaire (EEFQ)

Edinburgh Postpartum Depression Scale (EPDS)

European Quality of Life Five Dimension

Family Outcome Star

Fears of Compassion Scale

Generalised Anxiety Disorder Assessment (GAD-7)

Goal Based Outcomes (GBO)

Health of the Nation Outcomes Scales (HoNOS)

Hospital Anxiety and Depression Scale (HADS)

Kansas Parenting Satisfaction Scale (KPSS)

Karitane Parenting Scale

Keys to Interactive Parenting Scale (KIPS)

Levels of Adaptive Functioning (LOAF)

Maternal Antenatal Attachment Scale (MAAS)

Maternal Sensitivity Scale

Mother Object Relations Scales (MORS)

Maternal Postnatal Attachment Scale (MPAS)

Paternal Antenatal Attachment Scale (PAAS)

Parent Infant Interaction Observation Scale (PIIOS)

Parent Infant Relational Assessment Tool (PIRAT)

Parental interview

Parental Reflective Functioning Questionnaire

Parent-Infant Relationship Global Assessment Scale (PIR-GAS)

Parenting Stress Index

Parental Responsiveness Rating Scale (PaRRiS)

Patient Health Questionnaire (PHQ-9)

Postpartum Bonding Questionnaire (PBQ)

Self-criticism questionnaire

Service specific measure

Service User Feedback

Strengths and Difficulties Questionnaire (SDQ)

University of Idaho Survey of Parenting Practice (UISPP) (adapted for UK)

Video Interactive Guidance (VIG)

Work and Social Adjustment Scale (WSAS)

Working Model of the Child Interview (WMCI)

Appendix 3. Survey Respondents Job Role and Experience

3.1 Job role

Job role	Count
Psychotherapist	10
Clinical Psychologist	9
Children's Centre Child and Family Practitioner	3
Commissioning role	2
Consultant Psychotherapist	2
Early Years Practitioner	2
Social Worker	2
Specialist Health Visitor	2
Child and Adolescent Psychotherapist	1
Community Outreach Worker	1
Consultant Clinical Psychologist	1
Counselling Psychologist	1
Educational Psychologist	1
Family nurse partnership	1
Paediatrician	1
Parent and Infant Psychological Wellbeing Practitioner	1
Peer Support Worker	1
Perinatal Mental Health Specialist Lead	1
Project Manager	1
Research psychologist	1
Speech and Language Therapist	1
Trainee Clinical Psychologist	1
Trial manager (research)	1
VIG (Video Interaction Guidance) practitioner	1

3.2 Years of experience working with families of children under 5 years.

Years	Count
0-2 years	4
2-5 years	12
5-10 years	8
More than 10 years	22

3.3 Interventions used with families

Intervention	Number
Video Feedback Approaches (VIG or VIPP)	31
Parent Infant Psychotherapy (Dyadic)	19
Parent Infant Psychotherapy (Triadic)	17
Parent-baby groups	15
Solihull Approach	15
Watch, Wait and Wonder	13
Circle of Security	12
Incredible Years	11
Psychoanalytic therapy	10
Antenatal programmes	9
Peer support programmes	9
Infant massage	8
Mellow Parenting	6
Brazelton	6
Bespoke programmes	4
Watch me play	3
Baby Bonding - Compassion focused therapy	3
Parent only therapy	3
Creative therapy	2
Baby and Me	1
Parents as first teachers	1
Theraplay	1
Breastfeeding support	1
Mentalization based therapy	1
Family Nurse Partnership	1
Peep Learning Together	1