EVALUATION REPORT

NSPCC DART

Feasibility Study

The Policy Institute, King's College London

Beti Baraki, Hannah Piggott, Domenica Avila, Connie Woollen, Parnika Purwar, Irene Soriano Redondo and Suzanne Hall

March 2024

THE POLICY INSTITUTE

KING'S College LONDON

YOUTH ENDOWMENT FUND

About the Youth Endowment Fund

The Youth Endowment Fund (YEF) is a charity with a mission that matters. We exist to prevent children and young people from becoming involved in violence. We do this by finding out what works and building a movement to put this knowledge into practice.

Children and young people at risk of becoming involved in violence deserve services that give them the best chance of a positive future. To make sure that happens, we'll fund promising projects and then use the very best evaluation to find out what works. Just as we benefit from robust trials in medicine, young people deserve support grounded in the evidence. We'll build that knowledge through our various grant rounds and funding activity.

Just as important is understanding children and young people's lives. Through our Youth Advisory Board and national network of peer researchers, we'll ensure they influence our work and we understand and are addressing their needs. But none of this will make a difference if all we do is produce reports that stay on a shelf.

Together, we need to look at the evidence, agree on what works and then build a movement to make sure that young people get the very best support possible. Our strategy sets out how we'll do this. At its heart, it says that we will fund good work, find what works and work for change. You can read it here.

For more information about the YEF or this report, please contact:

Youth Endowment Fund

C/O Impetus

10 Queen Street Place

London

EC4R 1AG

www.youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

Registered Charity Number: 1185413

Contents

About the evaluator

The Policy Institute at King's College London was commissioned by the Youth Endowment Fund to undertake a feasibility study of Domestic Abuse, Recovering Together, starting in December 2022.

The Policy Institute at King's College London addresses complex policy and practice challenges with rigorous research, academic expertise and analysis focused on improving outcomes. We bring extensive expertise with impact and process evaluations, combining the rigour of academia with a pragmatic and programme-oriented approach to deliver robust and insightful evaluations of complex initiatives.

For further information about this evaluation, please contact Suzanne Hall (Principle Investigator) at suzanne.1.hall@kcl.ac.uk or Beti Baraki (Co-Investigator) at beti.baraki@kcl.ac.uk.

Executive summary

The project

Domestic Abuse, Recovering Together (DART) aims to support mothers who are victims of domestic abuse to develop their relationship with their child and support their child's recovery from domestic abuse. The intervention aims to address both the immediate and long-term impacts of domestic abuse on children, such as conduct problems, emotional distress and challenges with peer relationships. Typically, the mothers (and their children) are referred to DART from a range of agencies, including social care, schools, health and the voluntary sector. Developed and delivered by the National Society for the Prevention of Cruelty to Children (NSPCC), DART then offers a 10-week programme, where up to six families attend weekly group sessions, typically at an NSPCC Hub (or at the external "scale-up" sites that also offer the DART programme). Facilitated by trained DART practitioners, sessions last for two hours and use a range of activities designed to strengthen the mother–child relationship, promote communication about the abuse and support one another through recovery. The first hour of sessions is dedicated to shared activities between mothers and children (such as completing puzzles or drawing pictures). Children and mothers then participate in separate activities before coming back together for a fun activity or mindfulness session. To be selected for the programme, children must be aged between seven and 14 and have experienced domestic abuse, and the perpetrator must no longer be part of the household.

YEF funded a feasibility study of NSPCC DART. It aimed to ascertain how the programme is currently being delivered, explore the user experience and detail the barriers to DART's cohort reflecting the ethnic diversity of the communities it works in. It also asked whether an experimental or quasi-experimental impact evaluation of DART was practically possible and acceptable to stakeholders and how large such an evaluation would need to be. To explore these questions, the evaluator collected seven case studies, four at NSPCC sites and three at external scale-up sites (where they analysed project documentation, interviewed staff and observed DART planning sessions). They also conducted interviews with NSPCC staff, reviewed other stakeholders who were delivering services in the domestic abuse sector and analysed DART administrative delivery data. A total of 41 interviews were conducted: five with senior NSPCC staff, seven with referrers, and 29 with practitioners. In addition, three observations were carried out, each involving two practitioners. In total, 75 families accessed, or were due to access, DART between January 2021 and October 2023 as part of this feasibility study.

Key conclusions
DART is currently being delivered at NSPCC Hubs across the country, in addition to 25 external scale-up sites, by trained DART practitioners. Across these sites, there was considerable variation in delivery. Sites adapted the DART delivery manual to suit their needs, while staffing and financial constraints caused significant challenges to delivery. The intervention needs to establish clearer referral routes and guidance on eligibility criteria.
Eighty-four per cent of families that started the programme across the four NSPCC Hubs analysed in the feasibility study completed it (with families completing an average of 9.5 sessions out of 10). NSPCC practitioners and referral partners perceived that the programme provided significant benefits to families, including improving the mother–child relationship, communication, emotional management and self-esteem.
DART practitioners identified several barriers preventing the participation of families from racially minoritised communities, including cultural differences, low awareness of DART, the lack of ethnic diversity among practitioners and language barriers. Some sites have started engaging with local minority communities to address these barriers, such as organising community-specific events and providing information in various languages. However, more work is required.
To conduct an impact evaluation, significant challenges relating to sample size, the management of administrative data and outcome measures would need to be overcome.
Acceptability of an experimental or quasi-experimental methodology for evaluating DART varies among DART staff and referrers. Resource constraints and ethical concerns would need to be addressed before future evaluation.

Interpretation

DART is currently being delivered at NSPCC Hubs across the country, in addition to 25 external scale-up sites, by trained DART practitioners. Across these sites, there was considerable variation in delivery. For instance, while most practitioners found the DART manual to be valuable, some sites found it outdated and complex and adapted it to suit their needs. In addition, staffing and financial constraints caused significant challenges to delivery, and many sites were understaffed. The intervention also needs to establish clearer referral routes and eligibility criteria. Across several sites, practitioners used their professional judgement (in addition to the eligibility criteria) to select families, which created ambiguity within and across sites.

Eighty-four per cent of families that started the programme across the four NSPCC Hubs analysed in the feasibility study completed it (with families completing an average of 9.5 sessions out of 10). The most common referral route among these families was through local authorities (accounting for 36% of referrals). NSPCC practitioners and referral partners perceived that the programme provided significant benefits to families, including improving the mother-child relationship, communication, emotional management and self-esteem. The study did find variations in the way outcomes were measured across sites. Standardising outcome measures across sites and involving families in future research (e.g. pilot evaluations) is needed.

Most sites included in the study were in predominantly White areas (with only one area having a White population of less than 75%). Of the 75 families that served in the feasibility study, 85% self-identified as White. The remaining 15% were from racially minoritised communities, including Black Caribbean, Black British and British Asian communities. DART practitioners identified several barriers preventing the participation of families from racially minoritised communities, including cultural differences, low awareness of DART, the lack of ethnic diversity among practitioners and language barriers. Some sites have started engaging with local minority communities to address these barriers, such as organising community-specific events. Translators and interpreters have also been employed to ensure non-English speakers or those requiring sign language interpretation can participate. However, some DART sites shared that integrating translators into groups can be challenging, and the costs of translating materials can also be an additional barrier for sites. In future, the evaluator recommends partnering with well-established organisations with ties to diverse groups, such as minority-led domestic abuse organisations or specialist domestic abuse services, to improve outreach.

To conduct an impact evaluation, significant challenges relating to sample size, the management of administrative data and outcome measures would need to be overcome. A number of DART sites rely on limited and ad-hoc referrals, which makes it challenging to gather a sample size suitable for a randomised controlled trial (RCT) or quasi-experimental design (QED) evaluation due to insufficient numbers. The project would need to recruit three times the number of families it is currently recruiting. The level of data management is also relatively low across several sites due to capacity constraints. In addition, the study found that sites vary in the outcomes they measure. NSPCC consistently uses validated scales to measure two outcomes (self-esteem measured through the Rosenberg Self-Esteem Scale and attitudes and behaviours measured through the Strengths and Difficulties Questionnaire). In contrast, the other non-NSPCC-led sites show a lack of standardisation, relying on self-developed questionnaires and free-text feedback. These constraints would all need to be overcome before proceeding to impact evaluation.

Acceptability of an experimental or quasi-experimental methodology for evaluating DART varies among DART staff and referrers. Participants recognised the benefits of evaluations and research, including that they can help increase funding and referrals and improve service quality and the reputation of DART by proving its impact. A major concern of many practitioners was resource constraints, which need to be addressed. Some practitioners also expressed ethical concerns about RCTs and denying access to the programme for families in the control group. If proceeding with a QED, a large comparator group is necessary, and significant resources would need to be allocated to identifying a good source of a comparator group.

YEF is exploring options for further evaluation and considering what steps would need to be taken to make future impact evaluation successful.

Introduction

Context and rationale

Domestic violence, commonly referred to as domestic abuse, is a pervasive social problem affecting countless families. It often goes unreported to the police and, as such, is a hidden crime. While domestic abuse can impact individuals of any gender, evidence shows a clear gender disparity, with women facing significantly higher risks of enduring repeated and severe forms of abuse, including severe injuries (Walby & Allen, 2004; Walby & Towers, 2017) and being killed (ONS, 2020a; ONS, 2020b) compared to their male counterparts. Women are also more likely to experience higher levels of fear as well as coercive and controlling behaviours (Dobash & Dobash, 2004; Hester, 2013; Myhill, 2015; Myhill, 2017).

In the year ending March 2023, The Crime Survey for England and Wales (CSEW) estimated that 2.1 million people aged 16 years and over (1.4 million women and 751,000 men) experienced domestic abuse. However, due to underreporting of this crime, the actual number is likely higher. Around 73.5% of domestic abuse-related crimes involved female victims (ONS, 2023). For the same year, the CSEW found that a significantly higher proportion of people aged 16 years and over in the mixed and White ethnic groups experienced domestic abuse compared with those in the Asian or Asian British groups, and the proportion of women who experienced domestic abuse was almost double among White women (6.0%) compared with Black or Black British women (3.1%) and Asian or Asian British women (3.0%)(ONS, 2023). The reasons for these differences are complex and may be related to underreporting to the police among certain communities, cultural norms, lack of access to support services and incomplete data. It is important to note that, due to a survey error, the latest data shared by the ONS (up to March 2023) relies on only eight months of data collection (ONS, 2023). Caution should, therefore, be applied when interpreting these findings, given the potential impact of the reduced data collection period on the quality of the estimates. This caution also extends to data from the previous year (March 2022), as it was based on six months of data collection between October 2021 and March 2022, and lower response rates were likely impacted by the COVID-19 pandemic (ONS, 2022).

Domestic abuse and COVID-19

The pandemic had a significant impact on the nature of domestic violence, with lockdown measures and social isolation leaving families vulnerable to domestic abuse. During the pandemic, the CSEW found a 7% increase in police-recorded domestic abuse crimes between March and June 2020 (259,324 recorded offences) compared to the same period in 2019 (242,413 recorded offences) (ONS, 2020). However, CSEW acknowledged the difficulty of directly attributing this increase to the pandemic, considering the possibility of a gradual increase in domestic abuse crimes over recent years due to higher reporting. Despite the ambiguity in police recorded data, domestic abuse services experienced a significant surge in calls to their helplines and online platforms. For example, Refuge reported an increase in demand for its helplines, from 10,500 visits per month in the first three months of 2020 to around 73,595 per month between April 2020 and February 2021, an almost 700% increase, perhaps reflecting the severity, frequency and/or intensity of the abuse and lack of accessible, external support.

This discrepancy between police-recorded data and increased demand for support highlights the complex nature of domestic abuse and emphasises the critical role of domestic abuse services and interventions in providing a vital lifeline to victims of domestic abuse.

Domestic abuse and its impact on mothers and children

Additionally, domestic abuse also has multiple and varying impacts on mothers, their children and their relationships, and its effect on children requires a nuanced understanding. This is also particularly significant given the increase in domestic abuse cases during the pandemic (EIF, 2021; Walklate et al., 2022), while, simultaneously, there was a decrease in the number of young people being referred to domestic abuse services (Donagh, 2020).

While it is important to avoid characterising all children exposed to domestic abuse as passive victims, as vulnerabilities can vary (Callaghan et al., 2016: 399), it is equally important to understand how domestic abuse can negatively influence a child's development, health and wellbeing (Straus, Gelles & Smith, 1990). According to Women's Aid, around one in seven children and young people under the age of 18 will have lived with domestic abuse during their childhood, placing them at a greater risk of a wide range of negative outcomes. There is evidence to show long-term impacts on children, such as increased aggression (Meltzer et al., 2009) or the development of "conduct disorder symptoms" (Bowen, 2017: 97). Additionally, children who have witnessed domestic abuse might also be more likely to experience long-term mental health issues, including post-traumatic stress disorder (PTSD), anxiety (Edleson, 1999; Ionio & Mascheroni, 2021; Kilpatrick & Williams, 1998; Maikovich et al., 2008) and depression (Maikovich et al., 2008; Russell, Springer & Greenfield, 2010).

Researching domestic abuse support programmes for children is, therefore, relevant to the Youth Endowment Fund (YEF), given the organisation's aim to "prevent children and young people becoming involved in violence" (YEF, 2023).

In relation to the mothers who survive domestic abuse, the relationship with their children can also have an impact on their experiences of domestic abuse as well as their recovery.

The mother–child relationship may be strained during abuse if the child is abused as a method of controlling the mother (Kelly, 1994) or hurt when trying to protect the mother (Humphreys et al., 2008). At times, mothers view their children as an important motivation to leave a domestic abuse situation (Humbert et al., 2014: 366; Secco et al., 2016: 639), and mothers also report seeing their children as a source of strength (Javaherian, 2007: 55) and hope for recovery (Humbert et al., 2014: 366). Recovery can also be made more complex by motherhood (Carpiano, 2002: 445; Javaherian, 2007: 54), leading to increased anxiety and guilt during recovery about the impact of domestic abuse on their children (Carpiano, 2002; Javaherian, 2007). Motherhood can also lead to "emotional fatigue" for women worrying about "losing credibility with [their] children after excusing and/or covering up the abusive [behaviour] of the father" (Carpiano, 2002: 446).

Despite the evidence that the impact of domestic abuse is intertwined for mothers and children, support services for children and mothers have often been siloed (Humphreys, 2010). As a result, there seems to be a gap in support service that works jointly with the mother and child, with a primary focus on enhancing the mother–child relationship.

It is, therefore, essential to generate evidence of what works when it comes to supporting mothers and children who are victims of domestic abuse. To determine whether a programme causally improves participant outcomes, it is important to implement an experimental or quasi-experimental method when conducting impact evaluations.

Randomised control trials (RCTs) are considered the `gold standard` method to assess the impact of an intervention due to their capacity to create a valid control group that can mimic the behaviour of the treated group in the absence of the treatment. The key feature creating this comparability is random allocation, which allows researchers to attribute changes in relevant outcomes (if any) to the intervention. Randomisation, therefore, delivers causal estimations by isolating the effects that other factors may have on the outcome. For this reason, when interventions demonstrate evidence of promise, it is important to assess the feasibility of conducting an RCT or quasi-experimental method to provide independent and robust evidence of the effectiveness of the programme.

The role of feasibility studies

This is a feasibility study to assess DART's readiness for a full evaluation. Conducting a feasibility study is an important step to ascertain the readiness of the intervention for the piloting stage and is also a critical component within the MRC framework[1] for the development and evaluation of complex interventions. It provides a structured method to identify the factors that affect the delivery of a programme and determine methods through which these may be considered. Through this type of study, we can also assess the viability of resources that are necessary for implementing the interventions adequately. This may include exploring the ability to recruit and retain the required service providers and considering the extent to which data to measure desired outcomes can be collected, and stakeholder perspectives can be accounted for.

Intervention

Domestic Abuse, Recovering Together

Overview

Domestic Abuse, Recovering Together (DART) is a domestic abuse recovery programme that has a unique focus on supporting the mother–child relationship, in addition to supporting other aspects of mother/child wellbeing. It was developed in 2005 by the National Society for the Prevention of Cruelty to Children (NSPCC) and delivered within its own hubs or service centres.

DART is based on the "Talking to My Mum" research undertaken by the University of Warwick, which shows that children's outcomes are improved if the non-abusing parent is supported to take an active part in the child's recovery from domestic abuse.

During the programme, mothers and children participate in a range of activities designed to strengthen their relationship, promote communication about the abuse and support one another through recovery. Examples of activities range from completing tasks that require cooperation, such as puzzles or drawing a picture, to mindfulness activities.

[1] Craig, P.; Dieppe, P.; Macintyre, S.; Michie, S.; Nazareth, I.; Petticrew, M. Developing and evaluating complexinterventions: The new Medical Research Council guidance. BMJ 2008,337, a1655.

While DART was initially delivered within NSPCC Hubs until around 2015, it was subsequently selected as one of the services to be scaled up by the NSPCC. Since 2016, it has been made available through external "scale-up" sites. This meant that the NSPCC would support external service providers to adopt the programme. Currently, DART operates in 25 local authorities and voluntary organisations across England.

DART was reinstated again within the NSPCC for a limited period for an internal impact evaluation, which was published in 2016. Following this, DART has been brought back into NSPCC Hubs on a longer-term basis in response to the surge in domestic abuse cases during the COVID-19 pandemic and in light of a second internal impact study published in 2020.

The DART programme theory of change that was developed by the NSPCC can be found in **Appendix B**.

Who (recipients)

DART is provided to mothers and their children who are primarily assessed as victims of domestic abuse and have managed to separate from abusive male partners. Children must be aged between seven and 14. If a mother has more than one child, they are able to attend the programme multiple times with each child, but only one child is able to attend at a time.

Typically, the mothers (and their children) are referred to DART from a range of agencies, including social care, schools, health and the voluntary sector.

The suitability of families for the programme is assessed using various inclusion and exclusion criteria. The eligibility criteria set by the NSPCC are shown in Table 1 below.

Table 1: DART's inclusion and exclusion criteria

Inclusion	Exclusion
• Child aged 7–14 • Lived with domestic abuse experience • Assessed as harmed by this • Perpetrator no longer part of household	• Maternal inability to participate in group (for example, severe mental health issues) • Child inability to participate in group (for example, severe cognitive impairment or behavioural issues) • Child is known to have been subject to other forms of abuse, for example, sexual abuse that is unresolved and may require alternative intervention

Source: NSPCC (2021: 23)

What (materials used, procedures and who the providers are)

DART is delivered at NSPCC Hubs across the country, as well as 25 external scale-up sites (predominantly voluntary domestic abuse organisations and local authorities), by trained DART practitioners. Both internal (NSPCC Hubs) and external sites are provided with a programme manual, a day and a half of in-person training for DART practitioners, an information pack with a guide to implementing the service and ongoing telephone support.

How (modes of delivery)

All DART sessions are delivered in person. Each group work session is structured into two parts. During one part, the mothers and children come together in a group, and in the other part, they meet separately. The DART sessions are facilitated by a team of four facilitators or DART practitioners. The DART session lasts two hours, where the first hour is dedicated to shared activities for women and children; following a short break, the mothers and children are then divided into separate groups in different rooms. Two practitioners are expected to support each group when the sessions split. Each group then engages in activities separately, and they come together as one group for the final 20 minutes to conclude the session. Often, the session can conclude with a fun activity or mindfulness session that everyone participates in together.

Dosage (where and how much)

This study involved four out of a total of nine NSPCC Hubs that had been delivering DART and four out of a total of 25 scale-up sites, all located across England and Wales. One NSPCC Hub and one scale-up site were in partnership and delivered DART together.

The programme is 10 weeks long; up to six families attend each DART group, and sessions are held at either the NSPCC Hub, scale-up site or external delivery site (for example, a community centre) once a week for a morning or an afternoon.

In total, 75 families accessed, or were due to access, DART between January 2021 and October 2023 across the four NSPCC Hubs and four scale-up sites involved in this feasibility study.

Training programme

Organisations that are interested in providing DART at their organisation can have two of the four DART practitioners attend the one-and-a-half-day-long DART training course with the NSPCC or "trained trainers". In addition to this, they are also provided with the DART service delivery manual. This approach ensures that knowledge is retained, even in the event of staff turnover. Additionally, both the internal sites (NSPCC Hubs) and the external sites are provided with telephone support during the first year of delivery.

The manual offers an overview of DART, including a summary of evidence and the scale-up process, and directions for delivery. Practical arrangements for DART practitioners to follow (including timing of the sessions, location and planning), information on expectations of facilitators (including the skills and experience needed) and referral and assessment processes are outlined in this part of the manual.

What follows are session plans for each of the programme's 10 weeks and numerous suggested delivery resources. Session plans clearly define desired outcomes and resources needed and provide an outline of the session, including expected timeframes. The session plans outline activities for DART practitioners to use for the group as a whole and for the mothers' and children's groups when separated. Around 40 pages of the manual are devoted to example activities. These include worksheets aimed at helping groups describe their emotions and the stages of grief and loss, as well as relaxation exercises for mothers and children.

The 2021 manual ends with a "practice guidance" section, which reflects the most recent evidence base, and it is to be used by facilitators/practitioners in conjunction with the service delivery manual, which was

initially developed in the mid-2000s. This section of the manual provides alternative resources for use throughout the 10 sessions of DART. The practice guidance also includes an updated definition of domestic abuse and additional mindfulness activities.

Outcomes of the intervention

DART aims to address the immediate and long-term negative effects that children are likely to experience if they have been exposed to domestic abuse and, thus, improve educational attainment and reduce related presentations to health services by:

- Reducing the difficulties experienced by the child, such as conduct problems, emotional distress and issues with their peer relationships
- Increasing the self-esteem of the mothers and children
- Increasing the mothers' confidence in their parenting abilities

Evidence and rationale for DART

The NSPCC has been testing DART since 2010 and has refined and evidenced the effects of the programme through two internal evaluations.

Existing academic literature on domestic abuse interventions that seek to improve the mother–child relationship is, however, relatively limited. For example, a recent systematic review of intimate partner violence interventions (Trabold et al., 2020) includes only one study of a treatment programme in Sweden for mothers *and their children*. This study of dyadic therapy provided to a small group of diverse women and their children found that the women's PTSD symptoms were reduced for at least 12 months post-intervention (Grip et al., 2011).

More relevant to a consideration of DART are two systematic reviews of interventions that have some focus on "mothering" and of interventions focused on the recovery of mothers and children. Both describe limitations: "Much is still unknown about how different clinical and social care services might improve outcomes for female victims of [IPV] and their children" (Anderson and van Ee, 2018: 1), and the heterogeneity of interventions and limitations of the current research base (in terms of sample, measures, design and implementation) means that "it is not yet clear which...intervention components are most effective" (Austin et al., 2019: 498).

While the above reviews acknowledge the limitations in the existing research, there is some evidence that "studies implementing a combination of separate and joint working [by mothers and children] were seemingly more successful in improving [a wide] range of outcomes" (Anderson and van Ee, 2018: 17), including PTSD, child and parent self-esteem and children's social problems. Including play in joint work has also "been observed to be successful in improving child adjustment and mother-child interaction" (Anderson and van Ee, 2018: 17).

What's more, other available studies emphasise that many children do recover once they're safe from violence (Wolfe et al., 1986) and that the non-abusive parent plays a critical role in supporting a child's recovery (Humphreys et al., 2006). In terms of parental support, initial evaluations of programmes like Family Vision – "a 10-week life coaching programme for lone parents or carers" being piloted in Exeter – are positive (NIHR, 2019). Parents who participated in the programme, which also aimed to improve the parent–

child relationship, "reported an increase in confidence and feelings of control [and in] many cases, relationships with their children improved through parents' ability to better understand their child's behaviour and needs" (NIHR, 2019). A mother–child intervention focusing on group and play activities was also found to improve mothers' self-esteem and increase positive interactions between mothers and children (Austin et al., 2019: 508); this seven-week-long therapeutic group work programme in Stockport helped "to enhance the psychological well-being of the mothers and their young children and in promoting positive mother–child relationships" (Dodd, 2009: 34).

Indeed, DART has published its own findings (as a result of two internal evaluations). The first one, published in 2016, compared outcomes for DART delivered in NSPCC Hubs against a comparison group of mothers and children attending play therapy sessions at Women's Aid to assess if the programme helped to improve outcomes for mothers and children following domestic abuse. The second impact study, published in 2020, compared outcomes in NSPCC Hubs with external scale-up sites delivering DART. Both indicated positive outcomes for mothers and children who attended (Smith et al., 2015). For example, the majority of mothers (62%) reported substantially improved self-esteem and a significant majority of children who reported their mothers had failed to show them affection reported substantially improved relationships after taking part in DART (88%) (DART Manual, 2021: 10). Additionally, over half of the children with moderate or high behavioural and emotional issues reported substantially lower levels after attending a DART programme (DART Manual, 2021: 11). The NSPCC has also published information on DART delivery online.

The internal evaluations of DART provide evidence of promise about DART's potential effectiveness. This is an important step in evaluating an intervention. However, the existing evidence does not support making causal claims about the programme's impact. That is, based on the existing evidence, it is not possible to assert with certainty that DART improves the outcomes of the families that use the service. This limitation stems from the nature and the inherent limitations of the evaluations conducted themselves.

The reasons are outlined below:

1. Both evaluations were based on small sample sizes, which were insufficient to detect the causal effect of the intervention. In DART's first evaluation (2016), 88 mothers had completed measures in the treatment group and 18 in the "control" until the follow-up. In the second internal evaluation (2020), there were 67 mothers with matched data available in the treatment group, 28 in the first comparison group and 93 in the second (although corresponding children did not have Strengths and Difficulties Questionnaire [SDQ] data). These small sample sizes, added to the low quality of the assessed comparator groups (as explained below), imply that the internal evaluations had limited capacity to detect a causal effect and may instead be capturing a non-causal association. In such a scenario, any observed effect would not reflect the programme's impact; rather, it would likely show pre-existing differences between the treatment and the control groups. For instance, among participants on the waitlist in the control group, those who engaged first might have done so because they were in greater need. As such, any effect identified would be reflecting these existing differences rather than reflecting the impact of DART. Additionally, the first evaluation showed a high attrition rate, and an assessment of whether that attrition is asymmetric was not conducted. Therefore, it is possible that the participants who dropped out of DART were different than the ones who stayed. If those who completed the programme were more likely to experience improved outcomes, the effect estimator would be biased towards a positive result, and causality cannot be claimed.

2. The comparator groups assessed in the evaluations are not high-quality counterfactuals. Determining the causal effect of the programme relies on having a comparator group that can mimic what would have occurred with the treated group in the absence of the intervention. The two internal evaluations do not report balance checks between the treated and comparator groups. For instance, the second evaluation used as a comparator group participants that were asked to wait before receiving DART. The report mentions that this group was drawn from the initial participants, and there is a possibility that they were substantially different from participants referred at a later stage. In such a scenario, the positive effect found in the evaluation may be more a reflection of the pre-existing difference between groups rather than the causal effect of DART itself. Randomising the treatment assignment is necessary to guarantee that treatment and control groups are comparable. As for other control groups, they are drawn from other programmes or at different time horizons. This risks that these evaluations compare groups that were following different trajectories. Providing suggestive evidence of parallel trends across treated and comparator groups could mitigate concerns around this risk; however, these considerations are not addressed in the reports.
3. Long-term impact data on outcomes is unavailable. Therefore, even when considering a positive effect (which should not be claimed as per the limitations discussed above), there is no evidence on how long the effect lasts. This lack of information constrains what could be asserted in terms of value for money, for instance.

These limitations raise uncertainty about whether the observed changes in outcomes can be confidently attributed to the DART programme. This lack of causal evidence is important. It has been shown across a range of sectors that interventions that seem plausible may actually have no effect; on average, the proportion of interventions that have no effect is 80% (White, 2019). There is also evidence suggesting that plausible programmes can be harmful (Macintyre, 2011).

The evidence gaps in understanding effective interventions for mothers and their children, together with the limitations of DART's internal evaluations, highlight the importance of conducting a feasibility study of DART. These gaps and limitations will inform the research questions that guide the feasibility study and ensure the assessment of a number of aspects of DART's existing delivery methods, user experience, its effect on different sub-groups and methodological considerations.

While DART has evidence of promise, the lack of causal evidence regarding its impact on outcomes for families shows that the next step in understanding the effect of the DART programme would be to conduct an impact evaluation. However, conducting a feasibility study to assess the readiness of DART for such an evaluation is an important first step.

Research questions

This research aims to assess the feasibility of conducting an impact evaluation of the DART programme. It will consider the feasibility of utilising experimental or quasi-experimental methodologies to evaluate whether the DART programme produces positive outcomes for participants.

To fully answer this research aim, the study will address the following research questions:

1. How is the DART programme currently being delivered at different sites?

2. What is the user experience of the DART programme?
3. What are the barriers to DART's cohort reflecting the ethnic diversity of the communities it works in? What strategies could be used to address these barriers?
4. To what extent is an experimental or quasi-experimental methodology practically possible for an impact evaluation of the DART programme?
5. What sample size would the DART programme currently be able to provide for an impact evaluation, and are there credible routes to increase the potential sample size? What is the estimated sample size required to achieve adequate statistical power for a future impact evaluation of the DART programme?
6. To what extent would experimental or quasi-experimental methodologies be acceptable to referrers, NSPCC and DART staff?

The feasibility study plan is available on the YEF website.

Success criteria and/or targets

To assess the DART programme's readiness for a full evaluation, a set of criteria was developed through a collaborative process between the evaluation team, YEF and NSPCC. This process was guided by research and aligned with the programme's overarching objectives. The criteria aimed to identify the key aspects for a successful full evaluation and guide YEF's decision about whether to progress to a full evaluation of the DART programme. The feasibility study considered key progression criteria using a traffic light system to indicate stop, change or improve, and go criteria across four domains. The progression criteria are:

1. Methodology: whether an RCT or a quasi-experimental design (QED) is practically possible
2. Sample size for the proposed design
3. Acceptability of the design
4. Relationship between the evaluator and NSPCC and scale-up sites

Table 2: Evaluation success criteria

Evaluation domains	Progression criteria	Research questions
Methodology: whether an RCT or a QED design is practically possible In order to consider whether an RCT or a QED design is practically possible, we assessed a range of factors. We examined the current delivery of DART, referral pathways, consent processes and access to administrative data, among other areas.* Understanding the service-user experience and how DART is delivered allowed us to understand how an RCT or a QED would be able to fit into the current functioning of the DART programme. In the case of a QED, we also considered whether it is possible to construct a suitable comparison group and to check the identifying assumptions of the method (e.g. checking for parallel trends pre-intervention for a difference-in-differences design).	Either an RCT or a QED is practically possible, with at least 80% of the checklist items met and no major challenges identified. This indicates a high level of feasibility and minimal risks in conducting the study.	RQ1: how is the DART programme currently being delivered at different sites? RQ2: what is the user experience of the DART programme? RQ3: what are the barriers to DART's cohort reflecting the ethnic diversity of the communities it works in? What strategies could be used to address these barriers? RQ4: to what extent is an experimental or quasi-experimental methodology practically possible for an impact evaluation of the DART programme?
	Either an RCT or a QED is probably practically possible, but there will be challenges to overcome and some significant risks. Between 60–79% of the checklist items are met, indicating that additional effort and resources may be required to address the challenges and ensure a successful study.	
	An RCT or a QED is not practically possible, or there would be a significant risk the approach would fail. Less than 60% of the checklist items are met, indicating that substantial changes to the study design or resources would be needed to address the challenges and risks.	
Sample size for the proposed design This criterion considers if the currently available sample size (participants who are accessing the DART programme) is adequate for an experimental or QED design.	The sample size is large enough, achieves a statistical power of at least 80% and allows for generalisability of the results.	RQ3: what are the barriers to DART's cohort reflecting the ethnic diversity of the communities it works in? What strategies could be used to address these barriers? RQ5: what sample size would the DART programme currently be able to provide for an impact evaluation, and
	The sample size achieves a statistical power between 60% and 79% and provides limited generalisability, but it is still sufficient to provide meaningful	

	insights, particularly relating to identifying routes to potentially increase the sample size and enhance statistical power further.	are there credible routes to increase the potential sample size? What is the estimated sample size required to achieve adequate statistical power for a future impact evaluation of the DART programme?
	The sample size achieves a statistical power below 60%, limiting the ability to draw meaningful conclusions.	
Acceptability of the design This criterion considers whether an experimental or QED design would be acceptable to practitioners delivering the DART programme, to referrers who refer into the programme, and to the NSPCC and scale-up sites more broadly.	At least 90% of the stakeholders, including NSPCC, scale-up sites and referrers, and the NSPCC ethics committee, we speak to are open to exploring an RCT and/or a QED being conducted.	RQ6: to what extent would experimental or quasi-experimental methodologies be acceptable to referrers, NSPCC and DART staff?
	Seventy to 89% of stakeholders approve the design, with some concerns or reservations.	
	Less than 70% of stakeholders approve the design, posing a significant risk to the project's success.	
Working relationship between the evaluator and NSPCC This criterion relates to whether there is a constructive and open working relationship between the evaluator, the NSPCC and scale-up sites. We assessed whether the DART teams are motivated to assign time and resources to supporting a full experimental or QED design and whether there are sufficient resources available for NSPCC and the scale-up sites to drive future work forward.	All of the relevant stakeholders (NSPCC, scale-up sites and evaluator) have a strong working relationship and are committed to the research, and NSPCC and scale-up sites confirm the availability of sufficient resources to support a full experimental or QED design.	RQ6: to what extent would experimental or quasi-experimental methodologies be acceptable to referrers, NSPCC and DART staff?
	Most stakeholders, but not all, express commitment to the research and have a positive working relationship, and there is availability of some resources to support full experimental or QED design, but additional resources may be needed.	
	There is no commitment to one or more key stakeholder, or NSPCC and scale-up sites do not have the resources to support the research activities going forward.	

*NB: Practically possible was judged by considering (among other things):

- **Clear referral routes and eligibility criteria**: participants can be easily identified and recruited based on well-defined criteria. This involves a qualitative assessment of the clarity and accessibility of referral routes.

- **RCT randomisation**: it would be possible to include randomisation in the referral and delivery processes, allowing for the establishment of a control group without causing any ethical issues.
- **QED comparison group**: it would be possible to construct a comparison group and check identifying assumptions (including access to historical data).
- **Consistent programme delivery**: the DART programme is delivered in a consistent way across different settings compared to the manual, with any variations adopted only to better meet the needs/be inclusive of diverse groups.
- **Administrative data accessibility**: the necessary administrative data, including information on dosage and attrition, is collected and accessible.
- **Project delivery timelines**: the project delivery timelines align and are appropriate for the timelines of an impact study design.
- **Measurable outcomes**: outcomes of interest can be measured accurately and reliably.
- **Spillover effects**: the risk spillover effect that may influence the results is considered to be low.

Methods

This section provides an overview of the ethical review and data protection process and information on the project and evaluation team.

Ethical review

We submitted the project to the King's College London Social Sciences, Humanities and Law Research Ethics Subcommittee for low-risk ethical approval (reference number: LRS/DP-22/23-35596).

To carry out research at the NSPCC, the research also needed to be approved by the NSPCC's Research Ethics Committee. The interviews and case study research were approved via proportionate review. The observations required full ethical approval (reference number: R/23/233). The study was submitted to the Research Ethics Committee to receive approval prior to any fieldwork taking place.

Informed consent, participant information sheets and safeguarding

All participants (DART practitioners, senior staff members and referrers) were provided information sheets and the contact details of researchers at King's in case of any questions. Informed consent to participate was obtained from all participants prior to interviews. The information sheet clearly stated that participation was completely voluntary and that individuals' employers would not be made aware of participation or individual responses. We also asked participants via the information sheet and at the start of each interview not to discuss specific cases of those attending or being referred to DART but to focus on general themes instead.

The information sheet also advised participants of the process that would be followed in case of potential disclosures. It was made clear that we, as a research team, may have a legal responsibility to report information about safeguarding concerns (such as child or adult abuse). If such a case transpired, participants were informed that confidentiality may have to be suspended to ensure their own safety and the safety of others in order to contact the relevant support service.

No such disclosures were made during this research; however, participants were provided with the contact details for a number of relevant support organisations, given that the interviews had the potential to touch on distressing subjects. The participant information sheets are provided in the appendix for context.

If we had not received a signed consent form prior to the interview, we provided participants with a form before starting the interview. The consent forms we used were designed by King's College London's Ethics team and were adapted for use in the current project (for example, participants were only asked if they consented to being audio recorded rather than audio and video recorded).

We reiterated to participants at the start of each interview that confidentiality and anonymity would be maintained throughout the research process and in any outputs. At the end of each interview, we asked participants if they would prefer for anything they shared to be redacted from transcripts.

Data protection

All data was held according to the King's data protection policy and procedures. All data collection also adhered to ethical practices, ensuring the confidentiality of information shared and the secure handling of data in accordance with the General Data Protection Regulation (GDPR) and King's data protection policy.

Participant data was not transferred outside the EU. All team members are trained in King's approach, and data security was overseen by the Principal Investigator.

Access to individual files and folders was provided on a by-permission basis only, with higher restrictions put in place for files including sensitive or individual-level data sources. Rights to edit or access permissions to those files and folders was limited to personnel with a research need to access the data. The Principal Investigator controlled access to the folder and regularly reviewed who had access and if it was still required.

Under GDPR, our legal basis for handing personally identifiable data for research purposes was "task in the public interest", and the condition for processing special category data was "archiving, research and statistics". King's research privacy notice provides more detail on this here:

https://www.kcl.ac.uk/research/support/rgei/research-ethics/kings-college-london-statement-on-use-of-personal-data-in-research

Project team/stakeholders

Delivery team

The delivery team was made up of key staff at the NSPCC:

- Associate Head of Development: Sophie Bell
- Associate Head of Implementation: Helen Gazzola
- Implementation Relationship Manager: Wendy Pimblett
- Implementation Support Officer: Michelle Toal
- Assistant Director: Claire Crabb
- Quality and Sustainability Officer: Kitty Williams
- Business Support: Jamie Clark
- Senior Business Analyst: Kurt Coulter

The NSPCC Team was responsible for supporting the evaluation team to carry out the feasibility study. They worked with King's to identify and invite case study sites to participate in the study and arranged the signing of data-sharing agreements with the sites. The team also supported King's to recruit NSPCC staff for interviews and share administrative data.

Evaluation team

This project was delivered by the Policy Institute at King's College London. The team comprised Suzanne Hall, Director of Engagement (Principal Investigator); Hannah Piggott, Research Associate (Co-Investigator – Qualitative Lead, to March 2023); Beti Baraki (Co-Investigator – Qualitative Lead, March 2023 onwards); Susannah Hume, Director of Evaluation (Quantitative Lead); Doménica Avila (Quantitative Researcher); Connie Woollen (King's Talent Bank Qualitative Research Assistant); Parnika Purwar (Quantitative Researcher) and Irene Soriano Redondo (Qualitative Researcher).

Suzanne Hall acted as Principal Investigator for the project. She had oversight on all project activities, ensuring activities were delivered to time and cost, and provided quality assurance for research deliverables and outputs.

Hannah Piggott acted as Co-Investigator and day-to-day Project Manager on the project up to March 2023.

Beti Baraki took on the role of Co-Investigator and day-to-day Project Manager on the project from March 2023 to completion.

Susannah Hume provided oversight and guidance on the quasi-experimental and experimental research designs that the feasibility design addressed.

Doménica Avila was the quantitative researcher on the project and provided support throughout the evaluation, especially for administrative data analysis.

Parnika Purwar supported administrative data analysis for the final report.

The Policy Institute team also made use of King's Talent Bank (an internal King's service that facilitates the recruitment of research staff) to recruit a research assistant, Connie Woollen, who provided support with qualitative fieldwork, data management and analysis.

Irene Soriano Redondo also provided qualitative support for the final report.

This section provides an overview of the participant selection process, data collection and data analysis employed in the evaluation.

Site selection

Site identification, sampling and recruitment

At the launch of the feasibility study, we held an online workshop with senior and frontline NSPCC staff and YEF staff. This was a collaborative process between the evaluation team, YEF and NSPCC, where a recruitment approach for case study sites was agreed upon. Sites were sampled to capture the variety of experiences of the DART programme, including the type of organisation, demographics of the community in which it works and the size of the DART programme. Resourcing and acceptability were also taken into account during the decision-making process. This was particularly important for scale-up sites that had not previously been part of the evaluation work.

DART is delivered at NSPCC Hubs by NSPCC practitioners and externally at so-called "scale-up" sites. Scale-up sites tend to be domestic abuse support organisations or local authority teams. As such, our sample – or selection of case study sites – was planned to comprise both types of DART sites.

The NSPCC team was responsible for approaching and recruiting case study sites to be involved in the study. Initially, the plan was to recruit six case study sites (comprising both NSPCC Hubs and scale-up sites), with a plan to add an additional four sites if possible. However, due to capacity constraints of the organisations, only seven sites were recruited (four NSPCC Hubs and four scale-up sites). Once case study sites had been recruited, staff at each site liaised with the King's team to provide access to project documentation, recruit staff and referrers for the qualitative interviews and arrange observations. The NSPCC team also supported the recruitment of their colleagues for our qualitative interviews with senior NSPCC staff.

Settings and locations where data was collected

As mentioned above, this study's case study sites comprised a combination of NSPCC Hubs and scale-up sites to capture internal (NSPCC) and external (scale-up) delivery.

NSPCC Hubs: three hubs across England (in the Midlands, Merseyside, and North East, Yorkshire & the Humber) and one in Wales were included in the study. One English hub was working in partnership with a local domestic abuse support service provider to deliver DART. Three out of the four had delivered DART recently, while one had experienced various delays and was yet to deliver again.

Figure 1, below, compares the national rate of domestic abuse-related incidents and crimes recorded by the police to those reported in case study sites areas. Three of the four hubs included in the study were in areas with similar or higher levels of *reported* domestic abuse incidents and crimes.

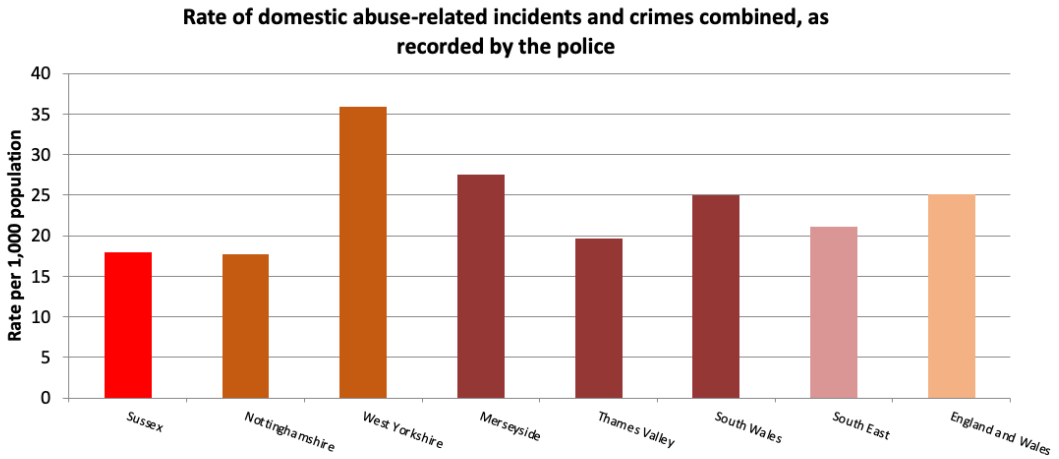
Among the hubs visited, one is housed in a purpose-built building, which is currently undergoing renovation to create additional child-friendly spaces for the on-site delivery of DART. It is worth noting that not all hubs deliver DART on-site. For example, one of the hubs in the study operates in partnership with a local domestic abuse organisation and does not offer on-site delivery. Additionally, another hub included in the study had not yet delivered DART due to delays in establishing the hub. This site was exploring the possibility of delivering DART in partnership with external partners and off-site locations.

Scale-up sites: twenty-five external sites across England and Wales hold an active DART license. We conducted research with four such scale-up sites, all of which were in England (in the Midlands, Merseyside, West Sussex, and Berkshire). Two were based in local authorities, one is a domestic abuse service support organisation and the final one operates in partnership with an NSPCC Hub.

The domestic abuse charity site included in the study used external spaces such as youth centres, children and family centres, and learning centres for programme delivery. However, staff at this site described challenges associated with external delivery sites, such as those related to access, as described below.

As per Figure 1, two of the four scale-up sites were in areas with lower-than-average levels of domestic abuse.

Figure 1: Rate of domestic abuse–related incidents and crimes combined, as recorded by the police



Source: ONS (2022b)

Description of participants involved in the feasibility study

Across all qualitative interviews, we sought to sample participants to represent the diversity of experiences of delivering and referring into the DART programme. For instance, we included both frontline, administrative and senior staff at the DART sites and referrers from a mixture of referral sources. We also aimed to represent the demographic diversity of staff and referrers, where possible. Our sample was, for the vast majority, White, reflecting the ethnic diversity of the organisations included in this study.

A total of 41 interviews were conducted in this study, which included five senior NSPCC staff, seven referrers, and 29 case study site staff.

Case study site staff (DART practitioners): due to the small size of the DART teams, a sampling frame was not used to recruit participants from within case study sites. At times, this resulted in interviewing the total population of DART practitioners at each site. One or two main contacts at NSPCC and scale-up sites were provided once the case study site has been selected. As such, between two and seven case study staff members per site were arranged through these contacts. This enabled the engagement of DART practitioners from across case study sites and of varying levels of seniority without overburdening staff members. The breakdown of participants per site is as follows:

- NSPCC Hub 1: 7
- NSPCC Hub 2: 7
- NSPCC Hub 3: 5
- NSPCC Hub 4: 1
- Scale-up site 1: 2
- Scale-up site 2: 4
- Scale-up site 3: 3

Capacity issues were an important consideration in this study. DART practitioners often had additional responsibilities, such as delivering other services in addition to DART and/or managing additional caseloads as children service practitioners or local authority early help workers. Practitioners at scale-up sites and partner organisations often had an additional role in domestic abuse services, working as, for example, an art therapist. DART practitioners who were involved in interviews did not only include those who directly facilitated DART sessions but also team leads, team managers and service managers who oversaw DART work, trained new DART staff or coordinated referrals.

A handful of practitioners across two hubs had delivered DART since its first introduction more than a decade ago, and one scale-up site participant had been involved in the first scale-up of DART in 2016. Most had become involved in DART, however, in the last two years or very recently, delivering DART for the first time within a few months of the interview.

Often, practitioners had worked directly in domestic abuse services (for a domestic abuse organisation or for the NSPCC in relation to DART) or indirectly (generally, as a social worker, for example, referring out to

domestic abuse services) prior to their current role. Some had been in social work or with the NSPCC for more than 35 years.

Referrers (recruited via case study sites): seven referrers were interviewed in total: one referrer associated with an NSPCC Hub and six referring into scale-up sites, each with varied levels of experience referring into DART and experience in the domestic abuse services sector.

Reflecting common referral pathways, three referrers we spoke to worked in early help or social work at a local authority, while another worked at a local primary school. In terms of experience with DART and the domestic abuse sector, one local authority referrer also delivered DART, while another had worked at a domestic abuse charity while training as a social worker. These local authority referrers had very varied experience referring into DART – between a couple of weeks and eight years. Another referrer we spoke to was working in a specific domestic abuse role within the scale-up site delivering DART.

Senior NSPCC staff: the senior NSPCC staff interviewed had varying levels of involvement in DART's day-to-day running, from close contact with hub(s) or scale-up sites to overseeing the team scaling DART up. This group of interviewees was chosen internally by the NSPCC. One senior staff member was closely associated with an NSPCC Hub, which was one of the case study sites, but was interviewed in their senior, strategic capacity rather than as part of the hub practitioner team.

Senior staff had varying levels of previous experience with DART delivery. However, one staff member had initially delivered DART. This allowed us to capture a broad range of views in relation to the management of DART at the NSPCC, the organisational aims for DART and the development of the programme at hubs and scale-up sites. Senior staff had been working on DART for between two and 11 years, and four of the five were female.

Data collection

The tables below outline the main methods we used for the feasibility study. The first covers the case study sites, and the second sets out the other methods we used. Details on the data collection methods are provided later in this section.

Table 3: Case study methods

Research methods		Data collection methods	Participants/data sources (type, number)	Data analysis methods	Research questions addressed
Case studies (6–10)	**Analysis of project documentation**	Review of documentation	Project documentation	Document review	1, 2, 4

	Staff interviews	Qualitative interviews	Up to seven per case study	Case and theme analysis; data triangulation	1, 2, 3, 4, 5, 6
	Referrer interviews	Qualitative interviews	Seven in total (one to two across four case study sites)	Case and theme analysis; data triangulation	3, 5, 6
	Observations	Observation	Three	Case and theme analysis; data triangulation	1, 2, 3, 6

Table 4: Non-case study methods

Research methods	Data collection methods	Participants/ data sources (type, number)	Data analysis methods	Research questions addressed
Interviews with senior NSPCC staff	Qualitative interviews	Five	Case and theme analysis; data triangulation	1, 3, 4, 5, 6
Environment scan	Review of stakeholders providing services in the domestic abuse sector	n/a	Environment scan	3, 5
Review of administrative data	Data provided by NSPCC and DART sites	DART programme data	Descriptive statistics	1, 3, 4, 5

Methods overview

Case studies: we carried out seven case studies in total. At each site, we conducted an analysis of project documentation and interviews with staff and referrers and, where possible, carried out observations of DART planning sessions. The case study work was carried out via in-person visits to four sites, where it was convenient for the case study site team, and online with three sites.

Qualitative interviews: all qualitative interviews were carried out using a topic guide (for guidance and examples, see Ritchie et al., 2003). Observations were carried out using an observation template (again, see Ritchie et al., 2003). These were designed in collaboration with YEF to address the feasibility study's research questions, and final versions of these were submitted to the NSPCC's proportionate review process.

Interviews were conducted by members of the evaluation team: Beti Baraki, Hannah Piggott and Connie Woollen. No members of the project delivery team were involved in data collection.

The following data collection methods were used:

Analysis of project documentation: project documents were reviewed to gain a deeper understanding of how each site is operating and to feed into the interview and focus group design. Relevant documents included project plans, quality assurance checklists or fidelity tools, adaptations the sites have made to the DART design, and public-facing documentation about the programme.

Staff interviews: up to seven qualitative semi-structured interviews (for a guide, see, for example, Ritchie and Lewis, 2003) were conducted with staff members at each case study site. At one site that was yet to run DART and recruit a full team, only one interview was conducted. Elsewhere, between two and seven staff of varied seniority were interviewed, from NSPCC assistant directors for the area and team leaders to long-term and more recently trained practitioners. Interviews covered staff members' experience of the DART programme, in particular, their sense of fidelity of delivery to the DART model and the practicalities and acceptability of different evaluation approaches. The latter included, for instance, considering how participants will be recruited into a trial, staff's views on randomisation, and how participants would be supported to complete the programme and engage with the evaluation measurement tools. Discussion with practitioners around the diversity of families supported on the programme and ways to improve the diversity of those supported was also covered in the interviews.

Referrer interviews: seven referrer interviews in total were conducted. Interviews included internal referrers operating within scale-up sites and partner organisations, as well as external referrers working in schools and local authorities. The relative ease with which referrers were recruited at scale-up sites could reflect the organisations' structure; internal referrers were easier to recruit via our contacts at the sites. We experienced significant challenges recruiting referrers at hubs, as reflected in our findings on barriers to referrals; generally, referrals are ad hoc in nature and close relationships between DART practitioners and referrers are few and far between. Moreover, many local authority referrers are time-constrained due to increasing workloads. This limited the ability of our site contacts to connect us to referrers within the evaluation timeframe.

These interviews covered what business as usual looks like in the absence of DART (and how this might change during an evaluation), what the wider system-level issues are that impact mothers and children accessing DART, referrers' current experiences of referring into the DART programme and referrers' views on different evaluation approaches. This included a discussion of randomisation and the feasibility of increasing referrals during an impact study.

Observations of DART planning sessions: fidelity to the DART design was assessed using a combination of interviews with DART practitioners as well as observations during the planning sessions. We observed three DART planning sessions across two sites (one hub and one scale-up site; at the former, we conducted two observations, as the final sessions of the course were being run separately for the two families involved). Ritchie and Lewis (2003) and Newing et al. (2010) provide detailed guides for observations, while McKenzie (2017) offers a practical account.

Assessing fidelity and adherence: fidelity to the DART design was assessed using a combination of interviews with DART practitioners and observations during sessions. During interviews, the extent to which each session adhered to the DART manual's prescribed content and structure was explored.

Additionally, practitioners were asked about their perspectives of the manual, any adaptions made and the reasons for this. Observations of planning sessions assessed the fidelity during actual programme delivery and compared practitioners' ways of working in real time to the manual's specifications. This enabled us to get a more detailed understanding of fidelity to the DART design and on-the-ground issues that could be present when attempting to use experimental or quasi-experimental approaches to assess the outcome of the DART programme.

We had planned to conduct observations at all NSPCC and scale-up sites included in the study. However, delays in the ethics process meant that a number of sites had finished delivering DART for the summer holidays by the time we received approval for observations.

Interviews with senior NSPCC staff (not necessarily affiliated to case study sites): five semi-structured interviews (Ritchie and Lewis, 2003) were conducted with senior staff members at the NSPCC who work on the DART programme. This helped clarify the experience of running DART at an organisational level and business as usual in the domestic abuse service sector. These interviews also helped identify key questions around the practicalities of different methodological designs, for instance, the process necessary to set up new sites (if these are needed to meet sample size requirements) and business as usual in areas without DART.

Environment scan and interviews with stakeholders in the domestic abuse sector: a review of stakeholders providing services in the domestic abuse sector similar to the DART programme was carried out. The aim was to conduct 8–10 semi-structured interviews (Ritchie and Lewis, 2003) with staff in these organisations. Numerous attempts to contact national organisations such as Refuge, Aanchal, and Solace Women's Aid were made. These interviews were planned to help understand how DART fits into the wider provision available for those who have experienced domestic abuse, how business as usual might change during the course of an evaluation and the wider system-level issues that impact on mothers and children accessing DART. It was hoped the discussions would help identify appropriate QED comparison groups. Unfortunately, it was not possible to secure interviews with key domestic abuse stakeholders. One national organisation shared that time constraints prevented them from being involved.

Review of administrative data: we carried out a review of administrative datasets from all sites currently providing the DART programme (not limited to the case study sites alone).

We worked with NSPCC and the scale-up sites to identify relevant administrative data that was used to consider the practicality of different methodologies. This aspect of the work was dependent on the data NSPCC and the scale-up sites collect, what they were able to share with us under GDPR, and the support of the NSPCC team and scale-up sites to access the data.

Changes to the feasibility study plan (protocol)

The feasibility study generally followed the feasibility study plan. However, some changes to the number of interviews planned were necessary. For example, referrer interviews were reduced from two planned at each site and 14 in total to seven in total across four sites due to limited established ties between sites and referrers and generally high workloads among social workers. High workloads across the domestic abuse services sector also meant we were unable to interview stakeholders following the environmental scan.

Some data collection was conducted online via Microsoft Teams, where convenient for staff. As such, we made fewer case study site visits than planned.

Analysis

Qualitative analysis of interviews and observations

Interviews were transcribed in full by a professional transcription service with a non-disclosure agreement in place with King's College London. A thematic framework matrix was developed in NVivo, and data was summarised into it (in line with "Framework Analysis"; see NatCen, 2020 and Ritchie and Lewis, 2003). A combined thematic framework with common themes that appeared across all interviews was developed, with sections added for additional themes found in only one or two interview types. This approach allowed data to be organised under descriptive themes while retaining the ability to view any individual's journey. During data management, the framework was reviewed by the research team to ensure its categories were discrete and exhaustive, and it was modified very minimally.

Once the qualitative data had been managed, it was analysed descriptively using a process of detection, categorisation and classification (again, following "Framework Analysis"; NatCen, 2020 and Ritchie and Lewis, 2003). This stage in the process involved team discussion to build a shared understanding of the data and to encourage internal challenge. Researchers conducted an explanatory analysis to look for linkages, associations and possible explanatory concepts to provide a rich understanding of how DART is delivered at different sites and what might impact the feasibility of using different methodologies to evaluate the programme.

When all qualitative data had been analysed, researchers came together to compare themes and explanations across the participant types.

Review of administrative data

Administrative data from the DART sites was cleaned and then analysed in STATA.

As described in detail in the sections below, descriptive statistics were used to understand the geographical spread of participants taking part in DART and their demographic characteristics (such as the range of ages and proportion of people belonging to different ethnic groups). We also sought to understand how participants engage with the DART programme and their pathways into and through DART.

In the data cleaning phase, missing values, outliers and inaccuracies in data entry were identified and accounted for. Given that all the sites had distinct data collection methods, we standardised the information and organised their respective datasets (when available) to establish common indicators. As part of the standardisation, a family indicator was developed to aid analyses at the family level. Creating this indicator is particularly important since the data from NSPCC was received at the individual level, with specific records for children and adults, while the other three sites delivered aggregated information at the family level. We therefore linked children and mothers to identify who received DART, thus harmonising datasets across all the sites.

To identify the attributes of the participants, frequency distributions were employed to illustrate how the data was distributed across categorical classifications, including their age, sex and ethnicity, for all families participating in the programme. Next, measures of central tendency were calculated, including the mean and median, for continuous variables such as the number of DART sessions attended by participants and the length of engagement within the programme. Furthermore, information on the heterogeneity of the data was provided by examining the interaction between the programme's participation and engagement indicators with variables such as sex, age and ethnicity.

When interpreting the findings, it is important to consider the limitations below:

Lack of consistency across sites: one of the main limitations to the analysis of the administrative data is the lack of consistency across all the sites. Given that each of the sites organises their data independently, we received extracts of information in different formats and with a range of details. This means that there are information gaps that have prevented us from including all the records in the analysis. For instance, some of the sites collected statistics on completion status, which could have been helpful in extrapolating the attrition rates in the intervention. However, it was difficult to incorporate this in the analysis as this information was not available for the four sites. Information on referral routes was also not available across all sites.

Another limitation in the analysis is related to the data received from one site, which was provided at an aggregate level. This means that individual records regarding the engagement of each family (mother and children) were not available. Instead, only data on the total number of families participating during a specific period was available.

Concerns about the accuracy of the data: during the analysis, certain inconsistencies in the data were found, which raised concerns about the accuracy and quality of the information. This was particularly highlighted around the number of DART sessions reported at the participant level since, in some cases, the recorded values exceeded 10, even after excluding sessions recorded as pre-assessment and assessment. This inconsistency could reflect data entry errors, especially when recording the type of activity conducted

with the participant. While the NSPCC have helped to clarify the data to enhance accuracy, it is reasonable to assume that there may be broader accuracy issues throughout the dataset. As a result, these accuracy concerns may affect the efficacy of the insights derived through the quantitative analysis, and therefore findings should be interpreted with caution.

Data triangulation

To facilitate data triangulation (see Flick et al., 2004; Leech and Onwuegbuzie, 2007), all data collection tools were designed around the common research questions and aligned with the progression criteria. This ensured that when analysing each strand of data, we could produce thematically aligned findings.

The team worked closely on all aspects of the project and were in regular contact throughout, allowing informal conversations to develop around common findings from the different methodological approaches.

To enhance data integration, we also held meetings and individual analysis by team members. Each team member scrutinised data independently and then came together to discuss findings and identify convergence and divergence between the various types of data to ensure a holistic understanding of the programme. These discussions led to decisions, clarifications and further analyses. Additionally, the team further reviewed the findings from each data source (both qualitative and quantitative) against the criteria, identifying whether the criteria were met or required further attention.

Findings

Participants

During this feasibility study, a total of 41 interviews were conducted: five with senior NSPCC staff, seven with referrers and 29 with practitioners from both NSPCC Hubs and scale-up sites. In addition, three observations were carried out, each involving two practitioners.

As outlined in the participant selection section above, we engaged with staff from case study sites who held varying positions of seniority; some had been involved with DART for up to a decade, while others had recently completed their training and started delivering in the months leading up to the interviews.

Interviews also included internal referrers operating within scale-up sites and partner organisations, as well as external referrers working in schools and local authorities. The senior NSPCC staff members interviewed held roles related to the daily management of DART or had a strategic oversight of the programme, though they were not directly involved in its delivery. Their experience with DART ranged from two to 11 years.

Intervention feasibility

Research question 1: how is the DART programme currently being delivered at different sites?

This section covers DART's delivery approach, introducing the sites included in the feasibility study and addressing variations in delivery methods, resources and capacity. In addition, DART's eligibility criteria and fidelity to the DART manual across sites are also covered.

Engagement with delivering DART

This study involved four out of a total of nine NSPCC Hubs that have been delivering DART. Their engagement with delivering DART has varied over time. Most notably, one hub was part of the initial development of DART in the mid-2000s. Other hubs have started delivering DART since the early 2010s. It is worth noting that all the hubs experienced interruptions in DART delivery before the onset of the COVID-19 pandemic.

Similarly, the establishment of scale-up sites also showed significant variations. For instance, at one local authority scale-up site, the introduction of DART was the result of a previous manager's encounter with DART a decade ago. This manager foresaw the potential for running DART at their location, believing it was the ideal site for such a programme. In contrast, another scale-up site, which was a domestic abuse service provider, got involved with DART in time for the 2016 scale-up study.

The final scale-up site involved in the study had already been delivering DART for approximately five years. Prior to delivering the service themselves, the local authority had been referring families to DART in a neighbouring local authority. The decision to shift to delivering DART within their own area was motivated by the desire to prevent local children from being placed on waiting lists to access support and also to avoid the costs and accessibility challenges associated with sending families elsewhere.

The process of setting up DART sites varied significantly, but all of the sites faced some challenges. This was particularly true for the hubs when DART was reintroduced in 2021. These variations were mainly due

to delays in gaining access to hub sites, challenges in securing referral pathways and the time needed to complete all the preparatory work required to launch DART effectively.

In one particular case, the delays in establishing the hub itself posed significant challenges; this site faced a reported delay of more than a year between accessing the site and finally establishing it as a hub. These delays had ripple effects on the development of referral pathways and the actual delivery of the DART programme. Consequently, as of the summer of 2023, this hub had not yet commenced the delivery of DART.

The substantial amount of work and time required to initiate DART delivery played a central role in many discussions surrounding the establishment of the DART programme. This involved activities such as training practitioners, setting up referral pathways and assembling the necessary resources and session plans.

DART implementation and delivery variations

DART practitioners at hubs and scale-up sites are provided with a DART service delivery manual. The DART manual offers guidance on delivering DART and provides numerous activity resources for the 10-week programme. The manual specifies that each DART group should have four practitioners, with two practitioners dedicated to supporting mothers and the remaining two to support the children when sessions are split into groups. Typically, groups consist of four to six families per year.

Sites described their current practices, with most running two to three DART programmes per year in line with school terms and leaving necessary breaks between groups for assessment. However, while some sites wanted to run two to three programmes a year, they faced challenges doing so due to low referral numbers, limited practitioner availability and suitable families withdrawing.

Two scale-up sites ran two groups concurrently, totalling between six to eight programmes per year. This approach was driven by high referral numbers or funding requirements. In the latter case, however, funding presented an additional pressure to find appropriate referrals:

"Obviously, there's pressure to fill the courses, but on the flip side of that, you want to fill the courses with the appropriate people because you want the best outcome for them". (DART Trainer)

This quote highlights a key challenge for programme coordinators: aiming to strike a balance between enrolling as many participants as possible potentially driven by funding pressures while also ensuring they are the right ones to ensure the best outcomes for families. This tension shows the importance of balancing accessibility with the need to provide tailored interventions that work for families.

Sites generally followed the NSPCC's recommendations regarding group compositions, which separate groups for younger and older children. Age considerations were also taken into account when planning groups, such as avoiding running groups for older children in September to prevent disruption for teenagers transitioning to secondary school.

Generally, both hubs and scale-up sites met the recommended thresholds of having four to six families per group. However, some hubs had to adapt the DART model significantly due to challenges with securing referrals. For example, one NSPCC Hub had to deliver the programme to just two families, with the families attending separately for the final weeks of the programme, mainly due one of the families missing

sessions. Another hub delivered to a single family, necessitating significant adaptations to the DART model. These findings highlight diversity in DART's implementation, the challenges faced and the need for adaptability to meet specific circumstances.

"We have had to adapt it because it's delivered to one child and one mother. So, I don't know any detail of that, but it couldn't possibly be the same as it would've been with a group of children and a group of women". (NSPCC Senior Staff Member)

Referral pathways: typical referrers to DART

Referrals to the DART programme were primarily ad hoc, with no well-established referral pathways, although several common referral sources were identified across the sites.

Local authorities, especially early help and social care teams, were the most frequent referrers to DART. Notably, scale-up sites based within local authorities often received "internal" referrals from other teams within their respective local authorities.

Schools were also commonly mentioned as a source of referrals, although the challenging funding environment in education was seen as hindering their ability to refer families to DART.

"I think education are in a place now where they struggle so much for resources that they don't have the time to make the referrals for DART". (NSPCC Senior Staff Member)

Referrals also occasionally came from other on-site support services at hubs and scale-up sites, some of which were specific to domestic abuse support. Additionally, GPs and court orders were infrequent sources of referrals.

Domestic abuse charities were frequently mentioned, but the nature of their relationships with DART sites varied. For instance, Women's Aid was reported to refer families to at least one site included in this study. However, there was a perception that the long-established history of Women's Aid was significantly limiting referrals. According to some participants, this limitation was attributed to Women's Aid's funding requirements, which prevented them from sharing referrals or waiting lists since their funding was tied to the number of cases they handled. This highlights a key trend in domestic abuse services, where significant funding constraints in the sector appeared to have an influence on service provision and relationships between the NSPCC and the domestic abuse services sector**.**

"Women's Aid has its own referral numbers to hit, and so they're not going to refer to us even if there's a waiting list because they will lose that referral". (Children's Services Practitioner/DART Practitioner)

"I don't think we see anybody as a competitor...I think the competition comes from the funding environment and the kind of uncertainty of funding for a lot of these smaller organisations". (NSPCC Senior Staff Member)

Referrers' relationship with DART sites

Referrers typically maintained contact with DART practitioners at hubs and scale-up sites throughout the DART programme. Child-in-need plans were particularly instrumental in enabling regular communication between local authority professionals and DART sites, often through core group meetings. According to

the Family Rights Group (2023), these core group meetings are responsible for ensuring that child protection plans are followed and regularly reviewed and that progress is monitored and adjusted if necessary.

Some referrers described how they could easily engage in phone conversations with DART teams to discuss potential referrals. DART staff often engaged in these discussions, carefully considering referrals to ensure they met the programme's eligibility criteria.

"I'll be honest, I find them very skilled...they do thrash it out with me as well, you know, in regards to what my thinking is [when referring a family]". (DART Referrer: Primary School Safeguarding Lead)

The study also found that referral pathways for DART were often unstable and difficult to establish. This was a challenge mentioned by all sites, particularly in the context of closures of key referral partners, often due to funding issues. The challenges faced by NSPCC Hubs in maintaining stable referral relationships were reflected in the study's difficulty in recruiting DART referrers for the feasibility study. Close relationships were both infrequent and unstable due to the volatile funding landscape within the domestic abuse services sector.

"Well, we did have a few [referral partners], but, unfortunately, they've – two of them have actually closed during this group; it's just funding…no funding". (DART Practitioner)

Furthermore, referrer–DART relationships had also been negatively impacted by the previous withdrawal of DART at short notice in 2018, following the internal impact evaluation. A number of participants shared that the intermittent nature of DART delivery within the hubs may have negatively eroded the trust of some referral partners who often relied on the service to refer families in need.

"I was actually a…local authority social worker, and I know that the people were quite annoyed that it had been stopped because it was such a powerful and effective service". (DART Practitioner)

Interaction between DART and domestic abuse services

The domestic abuse services landscape is diverse across England and Wales, resulting in varied interactions between DART sites and local domestic abuse services. These interactions ranged from formal partnerships at one hub to perceptions of negative interactions and competition with services at other sites.

Practitioners commonly referred mothers to domestic abuse recovery programmes or courses, such as Freedom or Gateway, before DART to ensure mothers were in the right place to support their children during their recovery in the DART programme. Additionally, the Rockpool Recovery Toolkit or "Connected" groups were also mentioned for post-DART support. However, it was noted that these programmes are designed for women only and are not available in all areas. Some participants recognised the challenges of this, noting that some families need additional support over and above DART.

"DART is a helpful tool...just a bit of the jigsaw puzzle. You know, it's not the be-all and end-all". (DART Practitioner and Trainer)

Sites also reported engaging with other domestic abuse services and local authority teams on behalf of the families they were working with to advocate for them. They reported that families could be overwhelmed

by the number of agencies they were expected to deal with during recovery, and so, their intervention here was to help the families manage this situation.

"We're getting referred people from professionals where [the] child is on a child protection plan and Mum, of course, just goes, 'Yes, I'll do DART' because they want to comply with the child protection plan. Mum is just completely overwhelmed with so many agencies, actions and services...so some of the work we have to do is advocate to get a better place…team around this family". (NSPCC Senior Staff Member)

Challenges and barriers in establishing referral pathways

Multiple hubs faced challenges in receiving sufficient referrals to deliver DART regularly. Barriers to establishing referral pathways included the domestic abuse funding landscape, the specifics of the DART programme itself and DART management.

Challenges in the domestic abuse funding landscape: the funding landscape for domestic abuse services is significantly limited and thus competitive and creates significant challenges for DART referrals. Cuts to local authority funding resulted in increased workloads for social care and early help professionals, who commonly referred families to DART. Sites reported inappropriate referrals being made by professionals who did not necessarily have capacity to fully understand DART's eligibility criteria and were taking a "scatter-gun" approach to support families in any way they could.

"Because local [authorities] are so desperate for families to engage in services, they'll just do a scatter-gun approach...and hope that one of them is suitable. It's...just a shining example of the current climate that we're in". (DART Practitioner)

Ultimately, this means that even if referrals are received, they're not always likely to progress to the group; instead, the referral to DART can come across as a box-ticking exercise. To complete all 10 weeks of DART, women need to be "really invested", but, in this site's experience, families referred by the local authority weren't always suitable to take part.

"Some women just are not in that right place when someone else has told them, 'You need to go to this'...Sometimes, it could be that it's a bit of a tick-box exercise, [and] a social worker could be suggesting it to tick that box". (DART Practitioner)

Additionally, the lengthy referral forms for DART deterred overworked referrers, particularly social workers, from making referrals. This reduced the number of referrals sites could receive, adding to the challenges.

"Particularly social workers...they're just as busy now...who have got 40-odd, 50 cases on the caseload; it takes half an hour, 45 minutes to fill that [referral] form out...and I do think that that in itself probably prevents us from getting more referrals than what we could do". (DART Lead)

Limited partnership due to competitive funding landscape: competition for funding between domestic abuse services also limited referrals. A well-established branch of Women's Aid was considered to be severely curtailing referrals from various public agencies to one hub in particular. The hope of developing partnerships with Women's Aid and sharing their waiting lists was hindered by the competitive funding landscape.

"If you're a woman who's suffered domestic abuse, unfortunately, you'll probably be referred to Women's Aid first". (Children's Services Practitioner/DART Practitioner)

"Women's Aid was our predominant domestic abuse service...and I suspect some of that is about how domestic abuse services are funded. So, although our service is different to their services, I think we would still be viewed as a competitor". (NSPCC Senior Staff Member)

Relatedly, the competition that exists between hubs and well-established local domestic abuse organisations manifested in friction over hub facilities and, as such, hubs' ability to sufficiently support families.

"Some of those organisations are so well established with their funding...that they have got a lot of things in place we don't have [like a creche], so that it'll be like, 'Oh, so what happens if they've got another child?' or, 'What, you take them out of school?' They're not saying, 'Well, we think that's a terrible idea', but they're asking questions in a certain way that leave me with the feeling that actually that is what they think". (NSPCC Hub Senior Staff Member)

Furthermore, the continuity of referral pathways faced disruptions caused by funding cuts, resulting in the closure of key referral partners. The loss of these key partners left hubs without established connections, severely affecting the regularity of referrals to certain hubs.

"We did have one called the Ruby Project, but that unfortunately lost its funding, and that is due to close in July". (DART Practitioner)

DART programme: in addition to the challenges created by the domestic abuse services funding landscape, there were also barriers associated with the DART programme itself. As discussed in more detail below in the "Eligibility interpretations" section, DART has a range of complex eligibility and suitability criteria that vary in use across and within sites. These criteria and a focus on the "readiness" of families led to a sense that **the right referrals** weren't necessarily coming in at the right time, leading to families being signposted to other services.

"I think it's been difficult getting the right people in at the right time. I think there's demand, but it's difficult because I know, unfortunately, we've had to turn a lot of families away, which feels really uncomfortable, but I think just on the referral criteria". (Children's Services Practitioner/DART Practitioner)

Additionally, the DART manual requires that groups be split between older and younger children, which is particularly challenging for sites that deliver DART less frequently**.** This often meant that families with children in one age category would need to be put on a waitlist or turned away until additional families with the right age composition were identified.

"The NSPCC recommend about a two-year gap, so we try and do that. Which means, though, if we've got two 14-year-olds, they can't come to the younger one, so we have to work out…can those younger ones wait? But the older ones can't. So, we do have to look at the referrals quite carefully". (Main DART Practitioner)

Finally, another aspect related to the DART programme itself was the timing of DART as a recovery service, which reportedly posed difficulties in establishing referral pathways. It was reported that DART was typically needed after professionals such as social workers or early help had already become involved with

families. As a result, it may not be straightforward to establish pathways, as the professionals might simply close the case once the family is no longer with the abusive partner.

"Often...local authority social workers, they weren't really working...with the mothers who would be suitable for DART, because once the domestic abuse perpetrator was gone, and they were no longer living with that person, they tended not to have the cases open". (NSPCC Hub Senior Staff Member)

Local authority teams like social work and early help are some of the most common referrers to DART, suggesting inconsistent relationships with local authority teams across sites.

Management of DART: DART has been delivered on and off at NSPCC sites for over a decade starting in 2005. DART was delivered by NSPCC sites until 2015, when the external scale-up process began and internal delivery began to withdraw. This partial withdrawal of DART aimed to create space for NSPCC sites to deliver other services as part of a "pipeline approach".

"We develop, we test, we learn, we redevelop, we test and then, once we've proven impact, we then scale. The idea has always been that [the] pipeline was like a sausage machine really...Once they're being scaled, we stop delivering them ourselves from our hubs to allow capacity for new services that are going through that development cycle to come in". (NSPCC Senior Staff Member)

As part of an internal impact study conducted in 2018, internal delivery was re-established but ended with the evaluation. Most recently, DART was brought back to NSPCC sites in 2021 in response to increasing levels of domestic abuse during COVID-19 lockdowns.

"I'm not sure had the pandemic [not] happened in the way it did we would necessarily have brought DART back". (NSPCC Senior Staff Member)

According to one hub, however, DART being pulled at short notice after the internal impact study created significant issues in re-establishing pathways when delivery began again in 2021. Practitioners described how some of the trust that was built may have eroded as a result and how one local authority, in particular, was "quite annoyed that it had been stopped because it was such a powerful and effective service". As a result, referrals were slow to pick up again.

"Because we'd had a service and stopped, and it does take a while for information to get out there and embed so that people know that there's another service". (Children's Services Practitioner/DART Practitioner and Trainer)

Devolved nations: also related to the management of DART are challenges particular to referral pathways in devolved nations. According to a Welsh hub, referral pathways with statutory services were hindered by the need for approval from devolved governments.

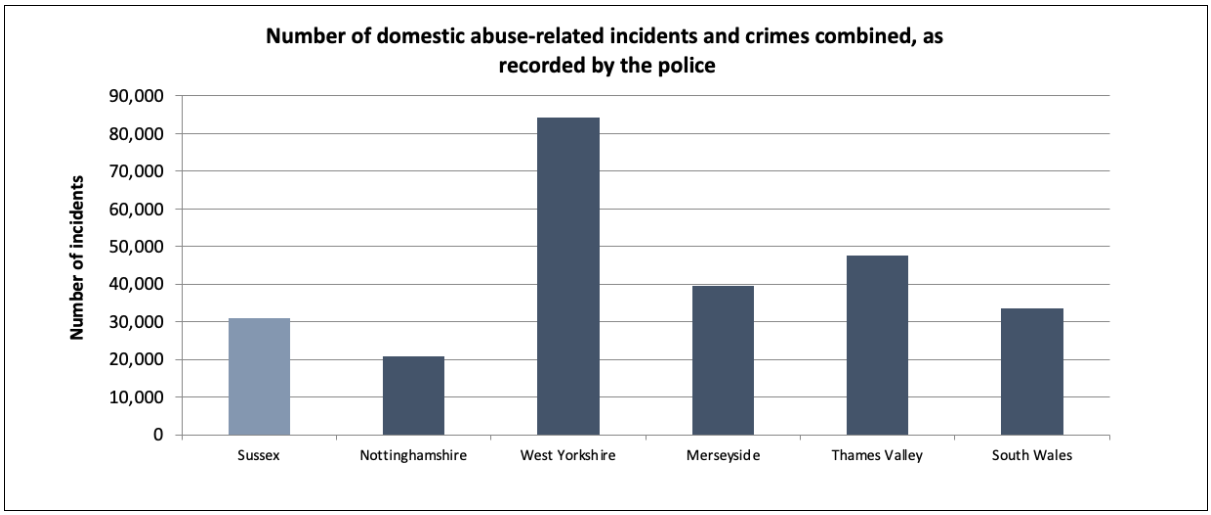
"CAFCASS [Children and Family Court Advisory and Support Service], I think, would be a good source of referrals, but they are not allowed to refer to services that aren't Welsh government approved. It needs somebody upstairs [at the NSPCC] to approach [the] Welsh government to say, 'Can they tick this box?' Because then CAFCASS can refer themselves". (Lead DART Practitioner)

Need and demand for service

The need for DART is evident in the high levels of domestic abuse reported in the areas where the study sites are located. However, the demand for DART does not always match the need. According to the participants, there is a generally perceived need for DART based on domestic abuse statistics, but this doesn't consistently translate into a high demand for the DART course itself.

Evaluating the need for service: to evaluate the need for the service, data is available up to March 2022 on domestic abuse incidents and crimes related to the police forces covering the study sites. (Note: two sites fall under one police force.) Five of the seven sites included in the study are in regions with domestic abuse rates higher than the national average in England. Figure 2 (in addition to Figure 1) also illustrates how these incidents and crimes compare to the national average in England. Notably, one site, yet to start DART delivery due to various delays, is located in West Yorkshire, where there are particularly high levels of domestic abuse.

Figure 2: Number of domestic abuse–related incidents and crimes combined, as recorded by the police



Source: ONS (2022b)

1Understanding demand for service: DART practitioners typically track demand by monitoring the referrals received. It is important to note that even when domestic abuse levels might be high across the board, predicting demand for DART on domestic abuse statistics alone is not sufficient. Planning stages should include efforts to understand the local need for domestic abuse recovery services while also identifying the presence of other available services in the area.

"I think, in hindsight, local scoping as to what other services were actually in the local area before making the decision to deliver may have been useful because it feels like it's such a needed service, but obviously needs are being met in other [ways]". (Children's Services Practitioner/DART Practitioner)

This scoping work is important because while there may be "interest" in or "need" for DART locally, reflecting high levels of domestic abuse, this need could be being met by other organisations in the area. For instance, in areas where Women's Aid is well established, demand for DART can be limited. In contrast, in regions where a site is one of very few available domestic abuse services locally, demand is much higher.

"We're really lucky because we are the only people in the vicinity that are able to offer a recovery programme for women and children". (DART Trainer)

This issue is further confused by the way in which different members of staff interpret and understand demand. For instance, the team leader at one site was clearly aware of the competition the hub was facing with another local domestic abuse organisation and how this reduces referrals. In contrast, a practitioner at the same site believed there would *always* be demand for DART due to the high domestic abuse statistics. However, this latter view assumes that everyone who experiences domestic abuse will need DART, which is not the case, as shown by the numbers accessing the service.

Eligibility interpretations

When discussing the eligibility of families, a distinction emerged at various sites. The distinction related to the official eligibility criteria set by the NSPCC and outlined in the DART manual, as shown in Table 1, and a set of additional criteria that rely on the professional judgement of practitioners. For the purposes of this report, we refer to this additional set of criteria as "suitability criteria". It is important to note that practitioners do not explicitly use the term "suitability criteria" but instead describe a range of additional factors that influence DART referrals alongside the criteria established by the NSPCC. While some inconsistency existed in eligibility criteria, there were significant differences and contradictions in "suitability criteria" within and across sites. The suitability of families for DART mainly depends on the "readiness" of a family, the mother's willingness to take part, the overall stability of the family unit and the potential for additional support to be provided to families that did not entirely meet the set requirements. The NSPCC expects professional judgement to be used in assessing and defining readiness for DART, leading to variations in how these criteria are interpreted across different sites. This is to ensure that a person-centred approach is used taking into account the individual circumstances of each family and that appropriate referrals are made. As a result, these "suitability criteria" can be ambiguous and open to interpretation.

"There are not a lot of reasons why you can't be part of DART. It's mainly about not being in the right place personally, which you don't get data from unless you start talking to someone". (NSPCC Children's Services Practitioner/DART Practitioner)

Challenges in interpreting eligibility criteria: the age of the child and relationship with the perpetrator were mentioned as part of the eligibility criteria by all sites (see Table 1 for a summary of DART inclusion and exclusion criteria). However, there were some discrepancies concerning the family's (mother and child) relationship with the perpetrator. For instance, practitioners at one site suggested that a family might be considered ineligible if they maintained any form of contact with the perpetrator, while another site suggested a family could be eligible as long as the mother was no longer involved in an intimate relationship with the abusive partner.

"Some of the referrers don't get that and just think it's about them no longer living together". (Children's Services Practitioner/DART Practitioner)

This aspect of DART eligibility is inherently complex due to the child/father relationship. In some cases, there may be a need for flexibility, particularly when contact with the father is required by a court order.

"Sometimes, there may be things in place like court orders for the children to see Dad, but we still feel that it would be beneficial for them to...have the course". (DART Service Manager)

The length of time since the family left the abusive relationship also played a role in these considerations, mainly due to the risk of post-relationship abuse. Assessments made on this criterion appeared to be highly individualised as different sites could set their own internal guidelines, but, ultimately, the family's specific circumstances influenced the decision.

"The NSPCC don't give guidelines on how long they should be out of a relationship. As a charity, we feel it should at least be three to six months. Again, it's all very individual. We could go and meet a woman and think, actually, you're so out of it, and you're ready. So that's kind of the only criteria, really". (DART Service Team Lead)

Understanding readiness and suitability criteria: the concept of "readiness" is not clear-cut. Some practitioners believe that only the family knows if they are ready, while others highlight the mother's emotional readiness to support their child.

"A lot of what DART is about [is] how full the mother's cup is at the time. Is taking on this service going to be too much? Are they not going to be a support for the child at this time, or is there room for it?" (Children's Services Practitioner/DART Practitioner)

Additionally, the willingness of mothers and children to participate in the programme, particularly in group activities, plays a crucial role in determining if it is the right time for the family to access DART. Mothers and children *wanting* to do the programme – and the group work specifically – was central to it being seen as the "right time" for the family to access DART. This was mentioned by one site in particular, which had experienced a string of families, who they considered not ready or motivated, being referred on a child protection plan. In such instances, the family might be signposted to other services until they become ready for DART.

That the families had undertaken previous work in relation to domestic abuse and its impact is also a factor considered in discussions of "readiness". However, there are significant variations within and across sites with regard to what counts when it comes to previous work. Some practitioners view previous domestic abuse recovery work as a prerequisite for DART, as they deem it might be more appropriate for the mother to have support in relation to domestic abuse and address any immediate issue (such as mental health support) before accessing DART in a group setting, while others are more flexible.

"Mum needs to do her own work first, so whether it's around mental health, whether it's around her understanding of domestic abuse...and it wouldn't be appropriate for them to be in that group setting". (DART Practitioner)

Ultimately, there appears to be no definitive or consistent guideline regarding mothers accessing previous domestic abuse work prior to being referred to DART. Further, given that domestic abuse courses like Freedom and Gateway are not available in all DART areas, it is not always possible for families to do this work prior to accessing DART.

"Readiness" is also dependent on stability and personal circumstances. Personal factors such as court or police involvement, a child's low school attendance, a parent's workload or having "too much on their plate" or "too much going on", which are not quantifiable, are also thought to impact a family's readiness

for DART. Additionally, stable housing is also a crucial consideration for "readiness"; DART might not be appropriate for families in the process of moving or residing in temporary accommodation as it would not be possible for them to complete the DART sessions.

"It has to be the right service at the right time for families". (Children's Services Practitioner/DART Practitioner and Trainer)

From speaking to practitioners, there was a sense that, ultimately, decisions were being made on an individual family level and on a case-by-case basis. Sites could offer long lists of complicating factors, and many did, while also concluding that extra support could be put in place to ensure a family could attend DART. Others would also assess a family and decide that, all things considered, they were not ready for DART, but due to a lack of other programmes in the area, this was all that was available to them. Ultimately, "readiness" is determined by the practitioner during the assessment process (as explained further below), and the practitioners use their professional judgement to come to a decision as to who should be able to attend or not.

"We will take a chance with a family. Sometimes, there isn't a guarantee that the family you've got, who have accessed support, are 100% ready, and sometimes, you think, 'Actually, this is all they've got at the moment, and I think that they could do this with some support.'" (Children's Services Practitioner/DART Practitioner and Trainer)

Assessment processes

The assessment stage is where practitioners assess the "suitability" of families after they are referred. Assessment processes and timeframes vary at hubs and scale-up sites.

These assessments typically involve multiple stages and administrative tasks. NSPCC Hubs conduct assessments in three stages: first with the mother, then with the child and then with the mother and child together. Children are only assessed when practitioners have a good sense that the mother is committed to taking part. On top of the DART assessment, hubs also carry out NSPCC assessments seven, 28 and 90 days after referral as part of their ongoing support provision.

Challenges exist regarding the consistency of assessment processes. Professional judgement plays a significant role in these assessments, as covered above, leading to potential inconsistencies in terms of who is deemed eligible.

Scale-up sites have created their own assessment forms following NSPCC advice, as required under the DART license. However, some inconsistencies between scale-up site assessment practice and ideal NSPCC practice seem to exist. One senior staff member described how hubs take up to eight hours to do assessments, while it's accepted for external sites to spend less time as long as they've assessed safety.

"We take up to eight hours doing our assessments. We appreciate they may not do that. But...that's absolutely fine, as long as they've assessed from a safety perspective and that the child and the mum is in the right position to take on DART". (NSPCC Senior Staff Member)

According to practitioners, assessment processes can be time consuming and act as a barrier to increasing the number of DART groups that run. This is particularly true at NSPCC Hubs, where DART assessments are carried out alongside NSPCC assessments, meaning that some of the assessment write-up is duplicated.

"I think four families is a good number to run...and, if I'm honest, I think the assessment process would be too long if there was any more". (DART Practitioner)

DART manual

Fidelity to manual: the DART manual is described in detail in the "Intervention" section above. The manual outlines the existing evidence base for DART and provides comprehensive plans for each of the 10 sessions, complete with clearly defined outcomes. Each session is two hours long and is facilitated by four DART practitioners. Mothers and children are split during the second hour to work on a particular theme or topic separately. Additionally, the manual includes resources for suggested activities, and the 2021 edition also includes updated practice guidance supported by more recent evidence (the DART manual was initially developed in 2005).

During the study, scale-up sites and NSPCC Hubs closely adhered to the 10-week structure of DART, following the weekly outcomes and key topics as laid out in the manual.

"DART is really specific, highly manualised…it is what it is across the 10 sessions". (NSPCC Senior Staff Member)

However, within this structure, practitioners are encouraged to use the manual flexibly to better cater to the unique needs of the families they serve. This enables sites to further develop the DART approach as they deliver more and gain experience.

"If they wanted to, [practitioners] could literally just pick up the manual and deliver it as it is. But they don't have to do that...they can bring in additional activities that mean it's right for those families...because all of the families they have will have different needs and respond to different activities better than others". (NSPCC Senior Staff Member)

Critiques of the manual: during the discussions about manual adaptations, critiques emerged regarding the need to adapt the manual. Some practitioners shared that they considered it outdated and not trauma-informed, stressing that many of the activities in the latest edition were developed in 2005 (NSPCC, 2021: 2) and that thinking has developed since then. For example, one practitioner raised concerns about the activities, in particular, the "Home Truths" video: a cartoon re-enactment of various domestic abuse situations. Parts of the video were considered inappropriate and dated.

"We have made a lot of adaptations to the manual. The service is very out of date; it's not very trauma-informed". (DART Practitioner)

"These families know what domestic abuse is; they've lived through it; we don't need to sit them there and re-traumatise them watching it on a video". (DART Practitioner)

For some practitioners from an NSPCC Hub, the manual was considered too complex, with too many activities for each session, which limited the time available for essential reflection. Additionally, certain

potentially "triggering" activities, such as the "Duluth wheel"[2], were not allotted sufficient time for reflection and discussion for families to process the content of the programme, causing concerns among some practitioners.

"We just literally felt so overwhelmed looking at this manual...I don't think that anybody should be going through a Duluth wheel and leaving it at that...that's something I would do [on] a one-to-one basis [because] a lot of this stuff is potentially triggering for these families". (Lead DART Practitioner)

Adapting the manual for group needs: practitioners often found it necessary to adapt the manual to accommodate the specific needs of the group they were working with. Most commonly, activities were adapted based on the child's age, requiring different materials, language and definitions for various age groups.

Furthermore, practitioners made adaptations to accommodate disabilities or learning needs, including deafness, autism, and anxiety. The manual was also adapted to cater to children's energy levels, likes and dislikes. For instance, a particularly energetic group might begin sessions with games to burn off energy, while other groups might prefer arts and crafts activities. That DART could be readily adapted to suit families' individual needs led to one practitioner describing it as inclusive.

Variations in resources and capacity to deliver DART

Staffing resources: staffing resources varied significantly across sites and were linked to discussions about the provision of the DART programme. The prevailing consensus among practitioners was that DART is a practitioner-heavy service, with practitioners setting aside one day a week to deliver each DART group and complete all associated administrative work. DART requires the availability of four to six practitioners each week per group, with four responsible for the programme delivery and two serving as backup in case of holidays or illness.

Outside of the group work, practitioners conduct assessments for upcoming DART programmes, make welfare calls to mothers between sessions, set up and pack up before and after the session (which can be significant given the resource-heavy nature of the programme), attend supervisions and maintain records of weekly "write-ups" for individuals and the group.

"There's a whole lot more to it, isn't there, than just the programme". (DART Lead and Practitioner)

DART was also not the only commitment that practitioners had. They often found themselves juggling DART alongside up to four other services and managing significant caseloads, which created significant challenges for them.

Working in partnership with other organisations appeared to alleviate staffing pressures at one hub. Overall, sites found meeting the staffing requirements challenging. Staff often worked part-time or had

[2] The Duluth wheel is also known as the 'power and control wheel'. It illustrates a number of different kinds of control used in domestic abuse to help survivors and perpetrators alike understand the domestic abuse behaviours they experienced or used (Domestic Abuse Intervention Programs, 2017).

additional caseloads to other service responsibilities, making it difficult to find the necessary flexibility for DART delivery. Additionally, the unavailability of DART training delayed the delivery process for new hires.

While volunteers could alleviate some pressure on staff, their impact was limited as they could not access internal systems to support associated administrative tasks.

Financial resources: financial resources also varied across sites. Funding is required to buy a DART license and to deliver the programme at scale-up sites. In cases where sites rely on sporadic grants for DART delivery, they might face the challenge of being unable to deliver for multiple years at a time until funding is secured.

Beyond licensing costs, financial resources needed for DART activities were considered a constraint across all sites. For example, one site shared they had collected all the resources they needed for the DART activities, and it amounted to "10 big plastic storage containers for each session". DART was described as "resource heavy", and such considerations were a significant factor in DART delivery for multiple hubs.

"…that was a lot of work; last summer was literally collating lists of all the resources we needed". (DART Practitioner)

Barriers and facilitators to expanding delivery: DART delivery is contingent on referrals, and the barriers created by a lack of referrals are discussed in the above sections. Staffing and financial constraints also create challenges for expanding DART delivery.

Practitioners described being understaffed while managing DART alongside other responsibilities. Increasing the number of sessions would require the recruitment and training of more staff, possibly doubling the size of the existing team up to a group of 12 (eight practitioners running two groups each week, with four for cover).

Other barriers to recruitment included the availability of DART training; one NSPCC Hub manager described having recruited a staff member, and two months later, DART training was yet to become available.

"[It] depends how often the NSPCC run the training, and that's a barrier in itself...I've got a practitioner that started in March, and I'd like to get her trained for DART, but the training's not even available to book on to; there's not dates or anything". (NSPCC Hub Team Manager)

Furthermore, any new DART practitioners hired would need existing knowledge of domestic abuse issues and the DART programme itself, particularly for related administrative tasks, including assessments.

Funding was frequently mentioned in the context of the need for more staff if DART provision was to be increased. This would be needed to cover the costs of hiring new staff and additional space and resources.

"More money. More money first of all to employ facilitators [practitioners] and to pay for everything else because all the resources cost money; the venues cost money...the equipment that we need [costs money]". (DART Trainer)

Research question 2: what is the user experience of the DART programme?

Descriptive analysis of the families served by DART

The following section provides insights into the families that participated in the DART programme during the evaluation period. The study did not directly engage with mothers and children as it was not part of the scope. However, it analysed nearly two years of DART programme data during which time it served a total of 75 families.

The section starts with an overview of the methodology used for data analysis. Following this, the limitations of the findings and key factors to keep in mind when interpreting the results are addressed. Following this, a snapshot of the families' demographics as well as information on the intervention's delivery are provided. These findings have informed the feasibility assessment for the impact evaluation, particularly in relation to the power calculation outlined in RQ5 below.

For the analysis, we included information from NSPCC and their four hubs (including the one involved in co-delivering DART with a scale-up site), as well as reports from the additional three scale-up sites.

Who is taking part in DART?

Over the course of the nearly two years analysed in this study, DART has served 75 families in total. Each of these families consists of a mother and a child. This overall figure covers all the NSPCC Hubs (four in total, including the one involved in co-delivering DART with a scale-up site) as well as the additional three scale-up sites mentioned above**.** Table 5 below summarises the share of families served per site.

Table 5: Share of families served per site

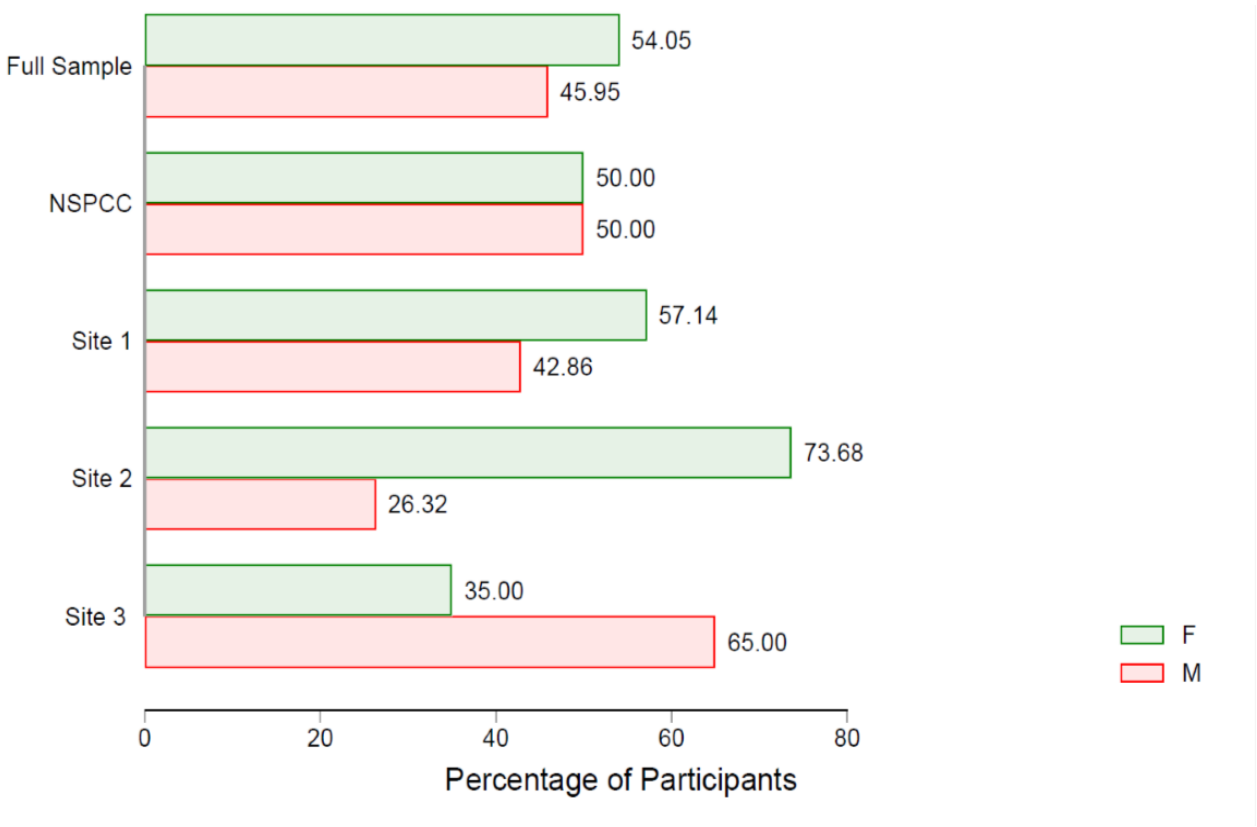
Sites	Served families	Percentage
NSPCC (four hubs in total)	15	20%
Scale-up site 1	21	28%
Scale-up site 2	19	25%
Scale-up site 3	20	27%
Total	75	100%

It is worth noting that while Table 5 above shows that NSPCC Hubs are serving fewer families (20%), there are reports accounting for around 80 adults who are currently in pre-assessment or assessment stages in order to determine their eligibility for DART. Some families have also been directed to other services based on their individual circumstances. Site 1 is the site serving the highest number of families (28%), although, overall, the distribution of families across sites is uniform.

The average age of the mother in the sample is 41 years, with the youngest mother being 29 years old and the oldest 50 years old. The average age of the child participants is 11 years, with the youngest being seven years old and the oldest 14.

Exploring further, we found that 54% of the children participating in DART are girls, as shown in Figure 3. However, the composition differs per site. For instance, while in Site 1, most of the children served are girls (57%), in Site 3, most of the children are boys (65%).

Figure 3: Children in DART, distribution by sex and site



When it comes to ethnicity, 15% of the sample are families that self-identify as non-White, and those served include Black Caribbean, Black British and British Asian. This distribution is consistent across NSPCC and the scale-up sites, but it is not possible to provide a more granular ethnic distribution due to concerns about the small numbers of the sample. We discuss further the diversity of the cohort engaged with DART as part of RQ3.

The most common referral pathways were also examined across the 75 families served by DART. While we did not have information on 21 of the families, it was evident that local authorities are the most commonly used pathway. Over a third (36%) of the families were referred by different services provided by the local authority, either the Integrated Front Door, children's social care services or early help. The second most commonly used route was self-referral. Therefore, if scaling up to a full RCT is to be considered, it would be essential to explore these two pathways in more detail to build the necessary sample size. Table 6 provides further details on referring organisations.

Table 6: DART referral pathways

Referral paths	Share of participants
Local authorities: Integrated Front Door/ children's social services/early help	36.0%
Third-party agency (e.g. Victim Support and Home-Start)	8.0%
Health services	2.7%
School	5.3%
Self-referral	14.7%
Other	5.3%
Not reported	28.0%

DART delivery snapshot and completion rates

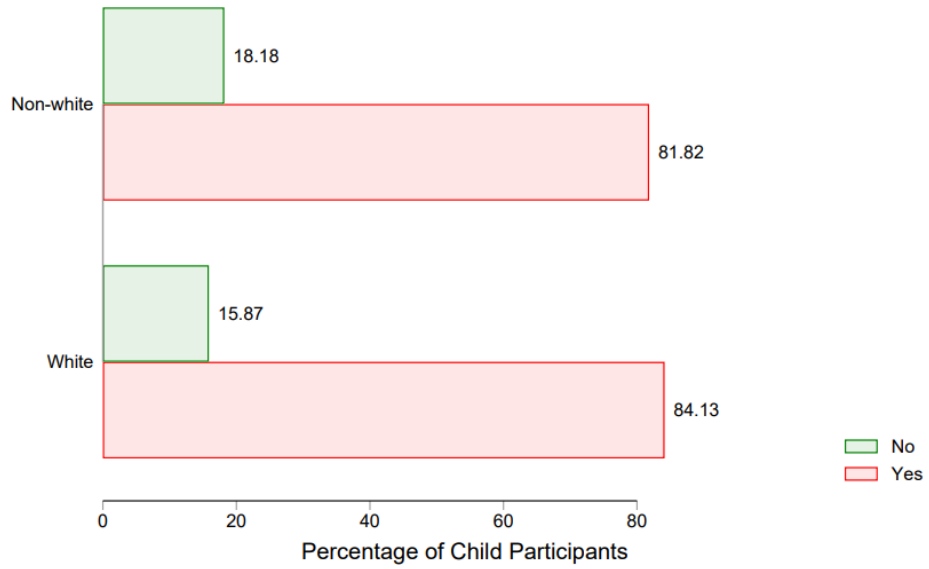
In summarising the delivery of the programme across all the sites, with information on families served since January 2021 and across the 75 families analysed, it is observed that, on average, families completed 9.5 sessions out of 10. This shows, overall, that the majority of families that start the DART programme tend to complete them**.** Table 7 provides a more detailed summary of completion rates for the programme. Specifically, only 16% of the served families dropped out or did not complete the 10 sessions during the period evaluated.

Table 7: DART completion rates

Completing status	Number of families	Percentage
Not completed	12	16%
Completed	63	84%
Total	75	100%

Figure 4**:** presents how the completion rate varies by ethnicity. Overall, there are minimal differences between both groups, with at least 80% of the families completing the programme.

Figure 4: DART completion rate by ethnicity



Findings from qualitative interviews suggest that attrition is low across most DART sites, and most families tend to finish the whole programme. Practitioners mentioned that some families might miss a session due to personal reasons, but they are then able to attend catch-up sessions to compensate. While essential for their progress through the programme, this adds to the resource constraints covered in the above sections.

"Most families do complete all of them, and from a children's point of view, I've never experienced a child not wanting to come. Obviously, they all start off quite nervous, but actually, they love it. They absolutely love it". (DART Practitioner)

One practitioner mentioned that, at their site, families are only allowed to skip two sessions, as missing more sessions can disrupt group activities. To avoid families not showing up, the site offers coffee mornings for mums to meet each other in advance and feel more comfortable. Practitioners identified other factors that can favour low attrition rates, such as offering lunch and snacks to participants, providing support with travel costs and sending reminders via text messages about upcoming sessions.

"We're halfway through our second programme, and we've had 100% engagement on both programmes so far. I think it's because we work in a person-centred way. I think it's because we can provide support with transport because I think that is a barrier at the moment with the cost-of-living increase". (DART Practitioner)

In the few instances where families dropped out of the programme, this was generally due to them moving to other areas or going through multiple difficulties, such as mental health problems or housing insecurity, making their participation at that time unfeasible. Only one practitioner mentioned that a family dropped out due to the programme itself as, according to the mother, the child was not ready to process the content of DART.

Finally, estimating completion rates is relevant as a proxy for expected attrition levels from the programme when conducting a full-scale RCT. Moreover, depending on when randomisation occurs, exploring take-up rates and number of participants at various stages of the referral process is important.

Although this information is only available from the NSPCC, we are providing some preliminary estimations that can offer an understanding of expected rates. Table 8 below summarises the number of participants at each stage of the process. Note that this covers participants beyond the 15 families identified as receiving the 10 DART sessions at NSPCC (excluding the scale-up sites), as we aim to provide insight into the overall pool of people NSPCC works with. Also, participants categorised as "in intervention" would be reflected in the pre-engagement (or at the referral point) and assessment phases, since those phases occur before the intervention is completed. It is also important to note that we included two intervention references in Table 8: "Intervention stage (DART and non-DART activities)" and "Intervention-DART". "Intervention DART" refers to the families engaged with the 10-DART sessions, while the numbers of the overall "intervention stage", as labelled by NSPCC, have also been included in the analysis to reflect the triage process happening at NSPCC, representing another potential point of attrition. It is important to acknowledge that not every family that has passed through the pre-engagement and assessment stages will necessarily receive DART sessions. This is an additional factor to take into account when considering the appropriateness of potential evaluation designs. The pre-engagement and assessment stages identify which participants can benefit the most from DART and which require other types of support or interventions.

Table 8: Intervention take-up rates (only from NSPCC)

Program stage	Freq.	Percentage
Pre-engagement or assessment	81	100%
Intervention stage (included DART and non-DART activities)	26	32%
Intervention – DART	15	19%

Perceived impact on mothers and children

DART practitioners and referrers widely agree that DART has consistently delivered highly beneficial outcomes for both mothers and their children. One of the most frequently observed outcomes, as noted by practitioners, is the improvement of the mother–child relationship as they undertake a recovery journey together. Providing families with a safe space to open up and express their feelings and experiences allows families to better understand each other and improve their communication and connection.

"We've had families that didn't have any physical contact, and the difference from one sibling to another was evident, and they've come to DART, and halfway through the group, the sessions, they're now cuddling…mother and child sitting on each other's lap and having a cuddle and laughing". (DART Practitioner)

Some practitioners mentioned that the programme has allowed mothers to better understand their children's behaviour, while it has also fostered empathy in children towards their mothers. This has helped both children and mothers realise that what they went through was not their fault and that it has happened to other people.

"They [a mother and her children] had really quite a good relationship, but by the end of DART, the child had a much better understanding of his experiences and maybe a better understanding of how things were for Mum as well". (DART Practitioner)

Practitioners also mentioned that, from their perspective, the programme appears to be effective in teaching participants about healthy relationships. Some of the key learnings include being able to identify unhealthy behaviours and improve their understanding of domestic abuse. These learnings can allow families to break the circle of abuse, especially when younger participants start to develop relationships outside their family circle, and further contribute to the improved relationship between mothers and children.

Improved emotional management is also a key positive outcome identified by the practitioners. By providing participants with tools to manage emotions when struggling, feeling angry or confused, as well as tools to better communicate their emotions, according to the practitioners, mothers and children seem to improve their behaviour towards other family members such as siblings and others outside the household.

"The little girl is now getting good behaviour awards at school because the school are seeing a massive, massive change in the way that she's presenting. She's not always reacting to situations. She's got those skills in herself to be able to deal with it". (DART Practitioner)

According to the practitioners, DART has also helped participants build resilience, self-confidence and self-esteem. For mothers, attending the programme is believed to be especially important in boosting empowerment and helping them rebuild their lives as their sense of guilt eases. This can allow mothers to get closure and distance themselves from abusive relationships. Feelings of empowerment and resilience can help mothers move forward and get new jobs and help young people re-engage in education.

"We've had teenagers that are not even going into school, but yet they have come on the group. And we've had teenagers that have gone back into full-time education after coming on the group, so we have some brilliant outcomes". (DART Practitioner)

By exploring families' needs and experiences during the assessments and sessions, practitioners have been able to identify other struggles that families might be experiencing and signpost them to other agencies to receive further support. Through DART, participants also developed a better understanding of support services available to them. Mothers were encouraged to communicate these issues to schools so they could also provide tailored support for their children. As a result of DART, some families have also left child protection plans.

"We would advise, historically, mums to let schools know what's been going on. That's very, very difficult for them to say, but at least by letting the school know, primary schools especially, we find can be very, very supportive". (DART Practitioner)

Although DART is regarded as a positive and enjoyable experience for most mothers and children who have experienced domestic abuse, practitioners have identified a few instances where DART has been less impactful. For instance, mothers experiencing severe trauma, in need of significant therapeutic support and who are not emotionally available for their children might not be able to achieve the above-mentioned positive outcomes.

Research question 3: what are the barriers to DART reflecting the diversity of the communities?

Diversity within DART cohorts

This section aims to explore the barriers to DART's cohort reflecting the ethnic diversity of the communities it works in and strategies already in use, or which could be used, to address these barriers.

Most sites included in the study are in predominantly White (British) areas. Only one area included in the study had a White population of less than 75%, according to the 2021 census, as shown in Table 9.

Table 9: Ethnicity census statistics for sites' local authorities

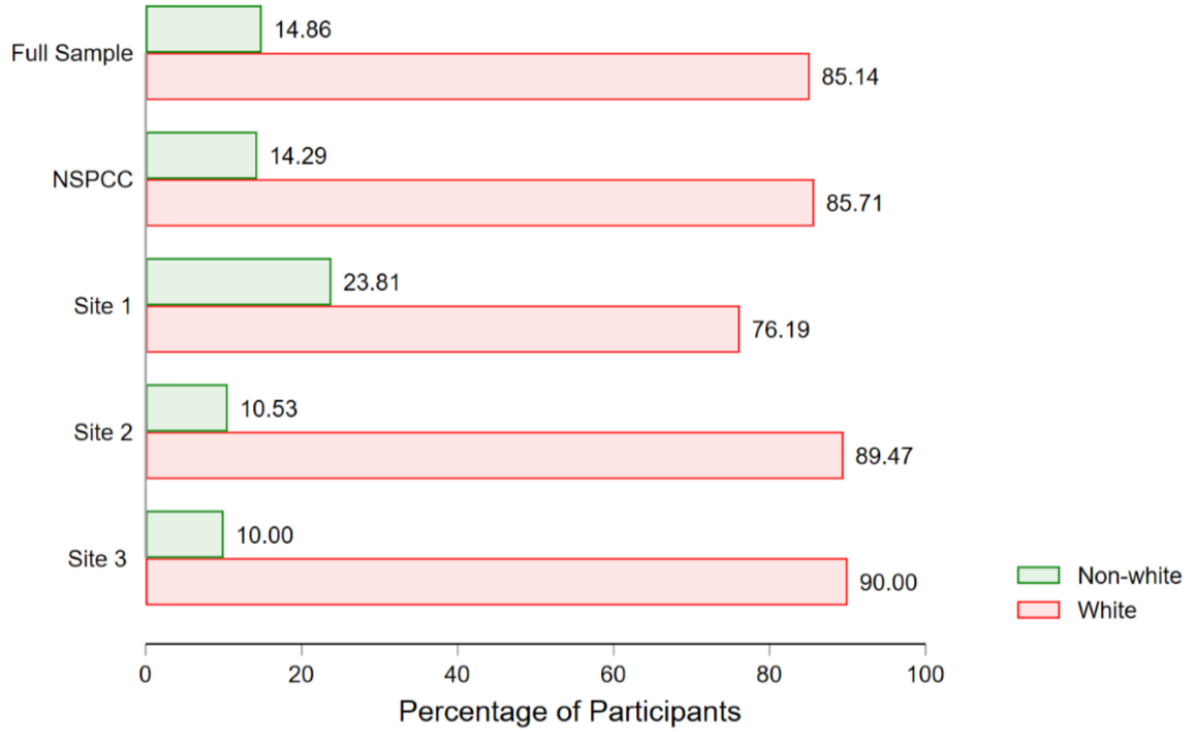
	2011		2021		Change	
	White	White British	White	White British	White	White British
Merseyside	98	96.6	96.5	93.6	1.5	3
Berkshire	90.6	84.9	86.1	77.8	4.5	7.1
Arun	97.1	88.7	95.8	88.6	1.3	0.1
Liverpool	88.9	84.8	84	77.3	4.9	7.5
Leeds	85.1	81.1	79	73.3	6.1	7.8
Nottingham	71.5	65.4	65.9	57.3	5.6	8.1
Cardiff	84.7	80.3	79.1	73.5	5.6	6.8

Sources: ONS (2012; 2022c)

This information is consistent with the findings from the descriptive analysis of the data provided by NSPCC and the scale-up sites. As presented in

Figure **5:**, 85% of the 75 families served by DART self-identify as White, and 15% of the sample are families that self-identify as non-White, including Black Caribbean, Black British and British Asian. This pattern is consistent across NSPCC and the scale-up sites, where site 1 is the one serving the highest number of non-White families (24%). This means that DART's cohort is reflective and representative of the overarching composition of the served areas and that the intervention is not oriented to benefit a particular ethnicity or community.

Figure 5: DART participants, distribution by ethnicity



Findings from interviews with DART facilitators and referrers suggest that DART participants are predominantly White British, and their first language is English, reflecting local statistics.

"[Regarding the diversity of referrals], I would say [referrals are] predominantly White British. Very rarely do our families of ethnicity come to us with these problems". (Referrer).

It's worth noting, however, that data is only available at the local authority level, and diversity *within* local authorities can vary significantly from one neighbourhood to the next, making it difficult to understand how well DART cohorts reflect local communities. Additionally, data is not collected by the NSPCC on local communities accessing hub services.

"I don't know [how well DART reflects local communities] is my honest answer...Because we don't have that in-depth discussion with them in terms of what they do and don't record. So, it's nothing that I could with real clarity and definitive kind of like, 'This is what happens.'" (NSPCC Senior Staff Member)

Elsewhere, sites reported increasing minority communities that are not captured in the overall census statistics. One site, for example, is in a local authority with the highest White British population of all areas covered in this research. However, staff were still concerned that DART was not reaching local minority communities and had employed an Eastern European worker to begin to connect to this growing group.

"We do have a very large Eastern European population, and they are probably underrepresented on the course". (Practitioner)

The focus of this site possibly reflects the in-migration of Eastern European migrants on a more local level than can be captured by census statistics. Concerns that minority communities are being missed were also raised elsewhere.

"It's possible that people are getting missed and there are unheard communities. It's very, very likely...definitely with people from different cultures". (Practitioner)

DART facilitators (practitioners) and referrers have identified different barriers that can hinder the participation of families from diverse backgrounds, even in areas where ethnic diversity is more prevalent. Barriers to participation include cultural differences impacting people's understanding of domestic abuse and their willingness to discuss and report domestic abuse incidents, lack of awareness and understanding of DART, as well as language barriers.

"One is a language barrier perhaps, and two is the fact that the cultural differences impact the awareness of what domestic abuse is". (DART Practitioner)

Other sites are beginning to engage local minority communities in their local areas. At one site, the local landscape is changing (as reflected in the relatively high level of change in the White British population between 2011 and 2021 in this area), and practitioners saw an increasing need to engage Sudanese, Syrian, Tamil, Afghan and Polish communities.

"Within our locality, the landscape is changing, if you like, because we've had lots of families that have moved in and have been fleeing war in Syria or Afghanistan". (DART Practitioner)

While staff were aware of the changing local population, it's difficult to say how well that is reflected in DART groups. While Tamil and Sudanese mothers had attended DART, a local authority social worker referring into DART commented that her caseload had still been predominantly White British, possibly reflecting the challenges faced by the domestic abuse services sector in engaging minority communities.

In terms of gender diversity, as DART is specifically designed for biological mothers and their children, the programme currently involves only women and their children.

Actions being taken in DART sites to address diversity

DART delivery staff acknowledge the importance of working to overcome barriers to participation and create a programme that is as inclusive as possible to ensure that people from different backgrounds can access their support.

"We can try and be as inclusive as we can....we're constantly thinking about how we can be inclusive to everybody in the best possible way, and some of it you do learn as you go." (DART Practitioner)

Some facilitators raised the importance of conducting initial assessments to understand the specific needs of families and discuss how the programme can be tailored to meet their needs. As outlined in the previous section on "Adapting the manual for group needs", examples of adjustments include adapting the language used in the sessions if younger participants are present or if there are participants with a disability. Other adjustments mentioned include being mindful of participants' cultural needs and beliefs or ensuring that there is space for wheelchairs in the room.

"We're very clear that any language issues and any disabilities [are] our problem, not theirs. It's about how we meet that need and about the solutions that we come up with". (DART Practitioner)

As mentioned above, a key barrier to participation is also the lack of understanding of DART and domestic abuse more generally. To address this, some DART sites are taking actions to reach out to their local communities, such as organising drop-in evenings for specific communities or donating copies of the Quran to mosques to use the opportunity to raise awareness of DART's work.

Addressing language barriers through translators and interpreters stands out as a key adaptation undertaken in some DART sites. Although English is the first language for most DART participants, some DART sites have reported employing translators and interpreters to ensure participants who do not speak English or require a sign language interpreter can take part in the programme. DART practitioners have stressed the importance of ensuring that interpreters and translators are integrated into the group and trusted by participants, given the sensitivity of the topics discussed.

Although efforts are being made to overcome language barriers, staff reported that there could be challenges to integrating translators and interpreters in groups, such as the difficulty of translating all materials, which can result in discussions being missed or a negative effect on group dynamics. Other challenges include resource constraints, such as the lack of facilities or capacity to employ them. The translation of material such as leaflets and consent forms – which is being done in some sites – was also considered costly.

"So, the difficulty we have is at [DART site], we don't have the facilities to have interpreters and things like that". (DART Practitioner)

These challenges have meant that not all sites are able to overcome the language barriers faced. For instance, a DART facilitator mentioned that a non-English speaking family was unable to join the programme as employing a translator was feared to negatively impact group dynamics, potentially interrupting interaction and increasing the length of the sessions.

"[Integrating a translator in the group] would double the length of time, and it would interrupt the interaction and the participation. A lot of it is talking, and a lot of it is games, and so it would be very challenging". (DART Practitioner)

Further suggestions to increase DART's diversity

One of the suggestions to increase DART's diversity proposed by several practitioners is to expand the eligibility criteria and update the manual (as currently the DART approach is based on evidence specific to women experiencing domestic abuse) to allow the participation of same-sex couples (e.g. females who had a female perpetrator), male survivors and children's non-biological mothers.

"I think it [DART] should encompass more, personally, because I think domestic abuse happens in every relationship regardless of gender…If we did branch out maybe into different genders…it might mean that there were more referrals for the service as well". (DART Practitioner)

In order to provide a safe and comfortable environment for women and children who have had male abusers – and, therefore, could be reluctant to join a group session with male participants – practitioners have suggested delivering gender-specific group sessions. Some practitioners mentioned that the language of the programme would not need changing as it already focuses on forming healthy relationships without focusing on gender.

"Potentially, some of these people [female participants] are absolutely petrified of maybe a male adult…I think it would be great for a male group". (DART Practitioner)

While some practitioners have suggested the possibility of expanding the programme to include male survivors and their children or same-sex couples, several others argued that there might not be sufficient demand for such services and that other specialist services might be better placed to offer support to male victims. As outlined in the section above on "Need and demand for service", despite the prevalence of domestic abuse in male-to-female relationships, the programme is already facing difficulties in receiving referrals for this specific group. Therefore, expanding the programme, for example, to same-sex couples might present even greater challenges in terms of referrals.

"We struggled to get enough families referred to the group for mothers. And we know that women are the higher number of victims of domestic abuse, so just trying to find enough fathers who have been victims of domestic abuse with children in a specific age range would be far too complicated". (Senior NSPCC Staff).

A few practitioners also suggested the creation of specific DART sessions for non-English speaking participants who make up a large proportion of the local community. For instance, facilitators from areas with large Polish or Arabic communities believe this option could allow DART to better reach and engage these communities, overcoming some of the existing participation barriers. For this to happen, DART sites would have to employ practitioners with different language skills.

"It would be great if, one day, we could deliver it all in Polish, and I can't see that being an issue if we had enough Polish-speaking people because any material can be adapted". (DART Practitioner)

Finally, improving DART's outreach could be key to improving participants' diversity. While some organisations described their efforts to ensure that local communities understand the key aims of the programme, many more have suggested that partnering with well-established organisations with closer

links to diverse groups (e.g. specialist domestic abuse services) could be an effective way to raise awareness of DART. The lack of diversity among DART practitioners – who are predominantly White – could also be a barrier to attracting diverse non-White participants. Therefore, partnering with minority-led organisations that have connections with their local communities could be a valuable approach to raise awareness of DART and overcome cultural and language barriers that can deter engagement.

Research question 4: to what extent is an experimental or quasi-experimental methodology for an impact evaluation of DART possible?

In this section, we discuss the status of DART on several requirements to conduct a robust impact evaluation. We start by outlining the advantages of experimental or quasi-experimental evaluations and the compelling evidence they create. We then introduce each of the different methodological approaches available for evaluation and synthesise their main requirements. After this, we assess to what extent DART complies with each of those requirements. We then summarise whether DART is ready to go into a full trial or QED and discuss potential routes for improvement. Recommendations for how to address any remaining challenges are discussed in the conclusions.

Causal analysis is a fundamental method used in impact evaluation that provides a rigorous medium to understand the "causes of effects". By isolating confounders, it allows not only reporting of the changes in outcomes that occurred over the course of the intervention but also recording of how much of the change may be attributed to the intervention itself. Additionally, it facilitates the understanding of the mechanisms driving the change, providing information about how the intervention would play if scaled up or transferred to a different setting.

Causal analysis is therefore essential for the following purposes:

- Determining the efficacy of a policy or intervention
- Providing insights about what steps can be taken to further augment the intervention's effectiveness
- Understanding the possibilities for scaling up an intervention or replicating it in other settings with a different population
- Gathering government support since it (causal analysis) informs government officials about the prospects for project success/failure, allowing them to make better decisions about how best to allocate public funds

Causal analysis works by constructing a counterfactual. The counterfactual refers to what the outcomes of beneficiaries would have been if they were not provided the treatment. The actual effect of the policy is then determined by comparing the counterfactual outcomes with the observed outcomes after the treatment. However, given that an individual cannot demonstrate both outcomes at the same time, the counterfactual outcome is directly unobservable. Therefore, counterfactual analysis is dependent on observing the outcomes of a control group, which does not receive the treatment and thus remains unaffected by the policy intervention. To obtain an unbiased estimate of the impact of the programme, it is necessary to ensure the control group is comparable to the treatment group.

RCTs, also known as experimental designs, are considered the "gold standard" in counterfactual analysis. By design, RCTs select a control group that is, on average, as similar as possible to the treatment group. However, in certain situations, conducting an RCT may not be feasible, thus requiring the use of an

alternative approach: the QED. This method differs from experimental design because it does not create a control group via the process of randomisation but instead exploits a source of randomness or exogeneity in the "real world" to construct a comparator group that, under certain conditions, is comparable to the treatment group.

The table below outlines the advantages of RCT as well as the different quasi-experimental methods used in counterfactual analysis, along with the requirements for providing an unbiased estimate of the treatment effect.

Table 10: Advantages and requirements of RCT and QED

Method	Advantages	Requirements
Randomised control trial	• Most rigorous tool for examining cause-effect relationships • Eliminates selection bias through random assignment of participants to treatment and control groups • Balances both observed and unobserved participant characteristics	• Selection of appropriate population type and outcomes of interest • Determination of appropriate sample size through power calculations • Random assignment into the treatment or control group • RCTs are more likely to involve more costly and time-consuming data collection methods. This type of data collection can also raise higher ethical concerns, especially when working with vulnerable participants or on sensitive individual-level outcome measures.[3] In contrast, quasi-experimental approaches are more likely to rely on

[3] RCTs are likely to rely more on primary data collection than other quasi-experimental approaches, especially in the domestic abuse sector. Primary data collection involves researchers having direct contact with vulnerable participants and potentially asking about information that could distress participants. In many instances, such research falls under "high-risk" ethical clearance and draws on a process of high scrutiny, potentially involving the development of safeguarding protocols to guarantee participant safety. A programme like DART requires primary data collection since there are no national surveys measuring the quality of the mother–children relationship, for instance. Moreover, even if such surveys exist, it is not always straightforward to map out from those surveys who is in the treatment and control groups due to data protection layers surrounding sensitive information. Quasi-experimental approaches are more likely to rely on secondary data, as, normally, comparisons could be drawn using a treated local authority and a control local authority, for instance. Also, since these methods require historical data, secondary sources are more likely to suit the needs of these designs better. In this case, participants will not need to be asked questions directly, and information will be accessed in secured environments or at an aggregate level.

		national surveys or large administrative datasets. • High-quality estimations are reliant on minimising attrition until the follow-up stage. • No contamination between the treatment and control groups
Difference-in-differences	• Provides an unbiased measure of the average treatment effect when randomisation is not possible • Controls for bias from unobserved variables that do not change over time	• Collection of data on outcomes in the group that receives the programme and the group that does not, both before and after the programme • Parallel trends assumption: differences between the treatment and control groups must remain constant over time to yield unbiased estimates. • There should be no potential confounding interventions happening at the time of the targeted intervention.
Matching and propensity score design	• Allows for making the control and treatment groups as comparable as possible to reduce bias, based on observable characteristics and under the assumption that the groups are similar on unobservable characteristics as well	• Needs extensive datasets with information on the observable characteristics of treated and non-treated units in the pre-intervention period • Reliant on the assumption that there exist no systematic differences in unobserved characteristics between the treatment units and the matched comparison units
	• Provides an unbiased estimate of the local treatment effect since individuals just above and below the cut-off	• A defined cut-off point is necessary since this would determine which population is eligible for the programme.

Regression discontinuity design	do not differ considerably • Doesn't require ex-ante randomisation • Prevents ethical issues of random assignment as it allows individuals who need treatment to be assigned the interventions	• The eligibility index must be continuous around the cut-off point. • The population of interest around the cut-off point must be very similar in observable and unobservable characteristics to yield unbiased estimates. • Individuals shouldn't move across the cut-off during the evaluation. • Requires a considerably high concentration of participants around the threshold to have enough power for estimations

Below, we discuss the evidence related to the requirements that the DART programme should meet for its evaluation in accordance with the methodologies specified above, as well as the key progression criteria identified in the project protocol (DART feasibility study plan). These criteria aim to facilitate YEF's decision about whether to progress DART to a full evaluation. Recommendations for how to address any issues appear in the conclusion. The elements assessed are listed below:

- Referral routes
- Eligibility
- Barriers to randomisation
- Constructing a comparison group
- Power calculations and sample size
- Measurable outcomes

Assessing referral routes

Typically, referrals are made to DART on an ad-hoc basis, from schools and local authority early help and social care teams or internally from other [domestic abuse] support services. Some sites reported a steady stream of referrals, sufficient to fill six to eight DART groups per year, while others described the inconsistency of referrals, receiving four one day and none for a period of time after that. Often, however, sites struggled to receive sufficient referrals from these sources. Relationships with local authorities vary across the board; practitioners and senior staff alike commented that social care teams might not always be appropriate as referral points as they are likely to have stopped working with suitable DART families by the time they are ready for a "recovery" service.

The ad-hoc and, at some sites, limited and inconsistent referrals reflect the funding landscape in domestic abuse support services. Most notably, organisations that once referred frequently to DART losing funding has left hubs looking for other referral pathways. Building these pathways is difficult as the high workloads of those working in domestic abuse support services and social work limit their ability to engage in the referral process. Significant funding constraints within the domestic abuse sector foster a competitive funding environment, limiting the relationships between DART sites and well-established domestic abuse organisations.

What this means for randomised control trials or quasi-experimental design approaches

The ad-hoc nature of referrals presents challenges for carrying out an impact evaluation. Referral pathways sufficient to form a sample for an RCT or a QED do not currently exist. Furthermore, the limited existence of reliable relationships with referrers and the ad-hoc and site-specific nature of many referral pathways means that there do not currently appear to be credible routes to increase the potential sample size.

Eligibility

The NSPCC sets inclusion criteria for DART, which stipulate, for example, that children attending must be aged between seven and 14 and that the perpetrator must have left the household. While eligibility criteria like age are followed categorically by all sites, there are some inconsistencies in the application of the requirements around the family's relationship with the perpetrator. In part, this results from the complexity of perpetrator/child relationships; court orders can require contact between the family and the perpetrator.

"Suitability", however, is much more ambiguous. While practitioners don't refer directly to suitability criteria, we use the term to refer to the various requirements set by practitioners across sites to gauge the "readiness" of a family. This depends on various factors, including housing stability and police or court involvement. It relies on professional judgement and, as such, varies across and within sites. DART's timing as a recovery service creates additional complexities and significantly shapes how many families of those reporting domestic abuse incidents could be suitable for DART; some may never want or be ready to access a domestic abuse recovery service.

There was a sense across multiple sites that DART is needed due to high levels of domestic abuse incidents reported in the local area – and this reflects local domestic abuse statistics, where the majority of sites had higher than average rates of reported incidents in England. However, suitability criteria appear to intervene, limiting the number of families that are able, ultimately, to access the service and providing a false sense of the size of the population that could benefit from DART. This presents challenges in predicting the sample size for an RCT.

The assessment process was described as time-consuming, especially at NSPCC Hubs, where NSPCC *and* DART assessments are carried out – both of which include at least three stages. Practitioners reported duplicating write-ups, and senior staff described how DART assessments were expected to last for eight hours. For practitioners delivering other services and working with significant caseloads, this is one factor limiting the possibility of increasing DART provision.

Scale-up sites are not required to conduct eight hours of assessments for DART; instead, they are expected to assess the safety and readiness of families.

Barriers to randomisation

The acceptability of randomisation is discussed in RQ**Error! Reference source not found.** below. In this section, we discuss when randomisation could occur in the recruitment process for DART. As explained above, randomisation would only need to be carried out if an RCT approach is being used.

During discussions with DART practitioners, two clear options emerged for when randomisation could be added to the DART process: at the referral point or after a family undertakes the pre-DART assessment. There was not a clear consensus option between these two.

The reasons for suggesting randomisation should occur at the referral point were that participants felt that they knew who was going to be eligible at that point anyway, that they did not think it would be acceptable to randomise someone once they had already gone through the assessment process or that the work done during the assessment process should be considered as part of the impact of the DART programme.

Where participants suggested that DART users should be randomised after assessment, it was because they did not think you would know if someone was eligible for DART until this process had been completed or because they thought randomising after assessment would reduce the overall wait time for those in the control group. From an analytical perspective, given the reported take-up and completion rates, randomisation should be conducted after the pre-engagement and assessment processes have been completed to maximise the chances of enough power[4] to detect the programme effect (if any).

This suggests there are two possible time points when randomisation could happen during an RCT. However, neither of them is without challenges.

Constructing a comparator group

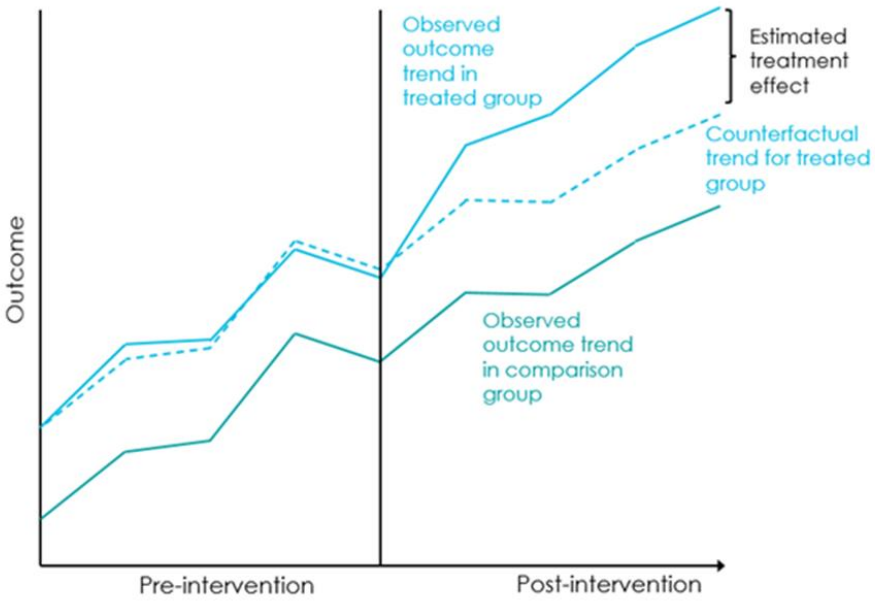
For a QED approach, we would need to construct a comparator group with different characteristics depending on the QED we are aiming to use. Below, we detail what building up a high-quality comparator group takes per method.

Difference-in-differences

Difference-in-differences (DID) is the main QED used to evaluate programmes when RCTs are not feasible. DID will compare changes in outcomes in a group that participated in DART against a group that did not across set time points. A key element of this method is the parallel trends assumption, which requires that the two groups are following a similar path for the outcomes of interest prior to the intervention. The parallel trends assumption is represented visually in Figure 6**:** below.

[4] Enough power refers to the capacity of the RCT design to find a statistically significant difference in outcomes across the treatment and control group provided this difference exists.

Figure 6: Parallel trends assumption



For this approach, rather than randomising participants to either condition, a comparator group would need to be identified that we expect to have the same trends on the key outcomes over the duration of the study. This assumption ensures that any divergences between the treated participants and the comparator group can be attributed to the intervention. If parallel trends exist, this method then cancels out any differences between the treated and comparator groups that existed before the intervention or that affect the groups during the evaluation period. Usually, this assumption would be tested by comparing the trends between the two groups before the evaluation period starts and making an argument that if the trends were similar before the intervention, they would have continued to be similar in the absence of the intervention. Checking parallel trends in this way is only possible where the outcomes are drawn from a longitudinal dataset, such as administrative data or a national-level survey with several-year waves before the intervention was implemented.

To apply this method in the context of DART, we will need the evaluation to focus on outcomes that are reported through longitudinal databases. Regarding the prevalence of domestic abuse rates, options could be using the CSEW or police registers.[5] Nevertheless, it is likely that we will not be able to match specific individuals receiving DART into these datasets, either because they do not record identification variables or because such matching is not permitted due to the sensitivity of the information. In that sense, although DID can be conducted at an individual or aggregated level, the feasible option would be to conduct an evaluation at an aggregated level (at the county level, for instance). This raises concerns regarding DART's sample sizes, since it will need to notably increase for the effect to be captured at an aggregate level. Another option would be focusing on other types of outcomes for the evaluation. Outcomes such as children's school attendance or performance are linkable at the individual level, using the National Pupil Database, for instance. Then, the question is around the likelihood of the effects of the intervention showing up in educational outcomes, according to the programme's model. DART's primary

[5] *Office for National Statistics*. (n.d.). *Domestic abuse in England and Wales overview.* Retrieved October 30, 2023. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/domesticabuseinenglandandwalesoverview/november2022

outcome focuses on the quality of the relationship between the mother and the child. As this is a soft outcome, conducting a DID could be very challenging. There are fewer surveys collecting longitudinal data on well-being or resilience at an aggregate level, which are commonly used for softer outcomes in the sector; therefore, DART's outcome would need to be measured by surveys, which is not available longitudinally.

Given the above, a DID design would not be practically possible for the DART programme.

Matched comparator group (propensity/exact matches)

Matched comparator evaluations compare participants who have received an intervention with other individuals who have not received an intervention but who are otherwise as similar as possible. Matching relies on a particular set of assumptions, the crucial one being that when we match a participant group to a comparator group so that all observable characteristics are similar on average, we have also been able to match them in all the unobservable characteristics. However, this assumption can be a risky assumption, as there may be unobservable factors that influence families' engagement with DART, such as self-esteem, which also affects the outcomes of interest.

There are a range of ways of conducting the matching procedure, including exact matching, coarsened exact matching and propensity score matching. Regardless of the specific matching approach, they all require there to be good levels of "common support" between the treated group and the potential comparator units. This means there needs to be sufficient information on the characteristics of both groups such that it is possible to identify good matches for treated individuals. Because of this, matching tends to require a large pool of potential comparator individuals that can be matched with a relatively small treatment group. Inadequate common support can lead to imbalanced groups, resulting in biased estimations of effect sizes.

A key challenge for this approach is recruiting a comparator group from which to find matches. As with the DID, this approach would be most feasible if the outcomes were confined to an administrative dataset. An alternative might be to find a broadly similar group of individuals and use a baseline survey to gather the data necessary to match participants. This would have cost implications due to the effort required to identify and recruit a high number of potential comparators that would be necessary to allow for high-quality matches.

During the study period, the potential of locating a comparator group and conducting surveys with them was explored with staff members. Suggested options to find a comparator group included working with local authority teams, partnering with other domestic abuse services, working with schools, looking at areas where DART is not delivered and working with the police. The domestic abuse services put forward were often very localised to the areas staff worked in. One organisation was mentioned repeatedly as a potential option: Women's Aid. However, staff also expressed concerns about finding a comparator group. They were unsure whether organisations would be legally able, or willing, to share data on the families they work with. It was also felt that getting the numbers needed for a comparator group would be very difficult, particularly as those who might have the contact details for a comparator group, e.g. local authorities and domestic abuse services, are overworked and unlikely to be able to prioritise an evaluation that would not directly benefit their organisation. This point was reinforced during the process of carrying out the feasibility study. We attempted to interview staff at national-level domestic abuse organisations,

in part to understand their ability and willingness to support an evaluator in identifying a comparator group. However, we were not able to recruit participants, largely due to their busy work schedules.

Matched difference-in-differences

The two approaches outlined above could be combined to "soften" the need for pre-intervention longitudinal data for an evaluation. It would, therefore, be possible to conduct a pre-post intervention survey covering both the treatment group and a poll of potential comparators. Based on the information collected, matching could be conducted to estimate a suitable comparator group. Since this group is, "on average", similar to the treatment group, it is possible to assume that both groups share parallel trends. While this is a feasible option, it's important to consider some additional requirements to the ones highlighted in the sections above. Since comparator participants will need to be surveyed at endline as well, an incentive scheme should be included to mitigate potential attrition and low response rates, as the comparator group will not have the same motivation to engage with the programme.

Given this, using a matched comparator and/or matched DID approach would be methodologically possible for the DART programme. However, there are challenges. Recruiting a comparator group would be difficult and would require significant resources. It would also likely be necessary to devote significant time to building relationships with a range of organisations that would have access to potential comparator participants. Given the under-resourced nature of the sector, funding would likely be required to support organisations in participating in the evaluation.

Regression discontinuity design

As discussed in Table 10, the main identification assumption for a regression discontinuity design is based on a clearly defined cut-off or threshold rule.[6] This approach compares individuals who are close to the threshold (or cutoff point) that determines eligibility for the programme. The key point is that being slightly above or below this cutoff point is almost like a random occurrence/assignment to the programme.. This similarity between the two groups makes it possible to attribute differences in their outcomes to the effect of the programme.

It's also crucial that individuals do not move back and forth across the threshold during the evaluation period for the regression discontinuity design (RDD) to be valid.

In the case of DART, the decision of who receives and does not receive the programme is not straightforward. While DART participants are selected based on the eligibility criteria and an assessment process, the selection process relies on a multifactorial assessment of the practitioners rather than a measurable indicator. Evaluations that use RDD typically work with clear-cut criteria such as students' attainment levels or means-tested benefits. We couldn't identify such a threshold in the delivery of DART.

While it is possible to develop a cut-off indicator with NSPCC, there are still concerns around sample size. Constructing a cut-off indicator could be possible if measurement instruments are introduced to DART's assessment process and the information is then aggregated into a single indicator. This indicator will then determine programme eligibility (who participates and does not participate in DART). This change would

[6] Lee, H., & Munk, T. (2008). *Using regression discontinuity design for program evaluation*. Proceedings of the Survey Research Methods Section, American Statistical Association

require a high willingness of the delivery organisation to measure and potentially modify their assessment process and enhance data management capacity for timely data collection.

Even if an indicator were constructed, an RDD would require a high density of families close enough to the threshold. Because of this data requirement, RDD evaluations are typically conducted on programmes rolled out at the national level. Given the relatively small numbers of families that are currently treated by DART, this approach wouldn't be feasible. This evaluation methodology does not suit DART's intervention design and delivery.

If a QED approach were taken to evaluate DART, the feasible approaches to constructing a comparator group would be a matched comparator or matched DID approach. However, both of these approaches would require a large comparator group to match from, which would be difficult to construct.

Outcomes measurement

There is high heterogeneity around the outcomes measured across the sites providing DART, both in terms of indicators measures and data quality. NSPCC, across all its hubs, collects information on two outcomes: self-esteem measured through the Rosenberg Self-Esteem Scale and attitudes and behaviours measured through the SDQ. Both instruments are validated and reliable tools for assessing such outcomes. While NSPCC adopted them as part of their normal practices after the previous non-experimental assessment they undertook, this doesn't seem to be the common practice across the other three scale-up sites. In fact, from the information collected, only NSPCC uses such scales. One of the scale-up sites uses a self-developed questionnaire where participants can express how they feel after DART in several aspects, including "free text" elements for participant feedback. Other sites use the Outcomes Star method, a questionnaire criticised due to its weak internal validity.[7] The third scale-up site uses a letters scheme for mothers to pass their experiences to the next DART cohort. The different sites providing DART, therefore, do not collect common outcome measures.

It is worth flagging that, as per DART's approach, the quality of the mother–child relationship is key to assessing the impact. Currently, this doesn't appear to be a variable measured across the sites. In the past, during the second internal impact evaluation, NSPCC used the Parent-Child Relationship Inventory for this purpose; however, it was discontinued after the evaluation was completed. Therefore, when thinking of conducting a full-scale evaluation, it would be necessary to identify a validated scale that can assess the relationship.

Additionally, the overall level of data management was relatively low. During the descriptive data analysis in RQ2, several concerns regarding the quality of the information provided were raised. It is also worth noting that the tasks involved in extracting information to inform this feasibility assessment appeared to represent a significant burden for the delivery organisations.

Readiness for trial

Below, we discuss the readiness for trial of the DART intervention.

[7] Sweet, D., Winter, K., Neeson, L., & Connolly, P. (2020). Assessing the reliability and validity of an outcomes star. *Journal of Children's Services*, *15*(3): 109–122. https://doi.org/10.1108/JCS-03-2020-0009

As discussed above, in order for a QED or an RCT method to be possible, there would need to be a significant increase in the potential sample size. This would mean that either DART sites currently delivering DART would need to increase the number of families they work with, or additional sites would need to start delivering DART.

As discussed above, DART is a resource-intensive intervention to deliver and to prepare for a trial; significant additional resources would need to be invested in delivering DART to reach the sample size required for an impact evaluation. Throughout the research, it became evident that establishing new sites and developing referral routes is a time-consuming process. Many of the NSPCC Hubs we spoke to indicated that they were still in the process of ramping up to full service delivery, with expectations of being able to deliver more DART sessions and attract more referrals going forward. This observation is significant, considering that the hubs had been operating for at least a year, highlighting the amount of time investment required for DART to successfully recruit and run sessions. As a result, it is reasonable to assume that a significant period of time would be needed before DART services are likely to have the numbers necessary to make an impact evaluation possible.

Progression criteria: methodology – whether a randomised control trial or a quasi-experimental design is practically possible

In summary, the practical feasibility of conducting a successful impact evaluation for DART depends on various aspects of DART that collectively determine whether an RCT or a QED is feasible. These aspects include sample and referral routes, a randomisation approach, comparator group construction, administrative data and outcome measures, as outlined in detail in the sections above.

Currently, the number of families accessing DART in the NSPCC Hubs and scale-up sites involved in this study is not sufficient to support a well-powered impact evaluation. The ad-hoc and, in some cases, limited referral routes into DART pose a challenge as there is no clear path to increase the sample size to a level required for a feasible impact evaluation.

For an RCT approach, there are two clear points when randomisation could take place, each with its own challenges regarding sample size and acceptability; these challenges are not insurmountable, but they would require careful planning and resource allocation.

For a QED approach, both a matched comparator and a matched DID design would be methodologically possible. However, both options need a large comparator group to match from. Considerable resources would need to be dedicated to identifying sources for a comparator group.

Regardless of the design option chosen, there is a need for improved administrative data collection and the identification of an appropriate outcome measure to take either of these approaches forward.

Overall, the current assessment suggests that, at this time, neither an RCT nor a QED design are practically possible. Substantial resources would be needed to increase referrals and delivery in order to meet the sample size requirements for a robust trial, and additional work would need to be carried out on administrative data collection and outcome measurements. While an RCT design might be somewhat more manageable to carry out than a QED approach, it is currently not deemed acceptable to staff members or referrers, as discussed in RQ6.

Sample size: sample sizes are discussed in detail as part of RQ5.

Research question 5: what sample size would the DART programme currently be able to provide for an impact evaluation?

Power calculations and sample size

The success of a robust evaluation relies on the power of the design to capture the programme´s effect (if any). The power of an evaluation design is influenced by several factors, including the size of the available sample, the outcomes of interest, the availability of sociodemographic information about participants, whether outcomes are measured at baseline, the level at which the intervention is implemented (individual, household or family), attrition and response rates, among others. In this section, we outline how these different factors interact with each other to simulate different power scenarios to measure the effect of DART.

Minimum detectable effect size

Minimum detectable effect size (MDES) measures the minimum size the programme´s effect should have in order to be detectable, based on the conditions of evaluation design. This means that if the MDES of an intervention is smaller than what the design has the power to capture, an evaluation will not be able to conclude that a programme improves outcomes, even if it actually does. Additionally, the effect sizes of the programme can vary depending on the outcome of interest. For instance, hard outcomes, such as the prevalence of domestic abuse or children´s school attendance, tend to show significant changes, although these changes might take a longer time to take place. On the other hand, outcomes such as improvements in the quality of the mother–child relationship might demonstrate the intervention's impact more quickly, but such changes are typically smaller. Therefore, the MDES depends on the primary outcome of interest.

Given that this is a feasibility study, measuring programme outcomes was out of scope. Therefore, we have conducted a rapid literature review to identify the effect sizes of similar interventions for relevant outcomes. As per DART´s approach, the primary outcome of interest is the quality of the relationship between the mother and the child. Additionally, effects on outcomes such as self-esteem and behavioural attitudes, which are measured through the SDQ, were considered. Therefore, the focus of the review was on meta-analyses and systematic reviews that included RCTs or quasi-experiments measuring these outcomes.

Interventions like DART report effects that typically fall within the range of small and medium sizes mostly,[8] particularly when they are low intensity. This is not surprising, given that these are soft outcomes and self-reported through surveys. Based on the findings, we will consider two possible effect size scenarios for our power calculations: 0.2 and 0.4 Cohen´s standard deviations. It is worth mentioning that

8 Anderson, K., & van Ee, E. (2018). Mothers and children exposed to intimate partner violence: a review of treatment interventions. *International Journal of Environmental Research and Public Health*, *15*(9). https://doi.org/10.3390/ijerph15091955

Howarth, E., Moore, T. H., Welton, N. J., Lewis, N., Stanley, N., MacMillan, H., Shaw, A., Hester, M., Bryden, P., & Feder, G. (2016). IMPRoving Outcomes for children exposed to domestic ViolencE (IMPROVE): an evidence synthesis. *Public Health Research*, *4*(10): 1–342. https://doi.org/10.3310/phr04100

Kiani, Z., Simbar, M., Fakari, F. R., Kazemi, S., Ghasemi, V., Azimi, N., Mokhtariyan, T., & Bazzazian, S. (2021). A systematic review: empowerment interventions to reduce domestic violence? *Aggression and Violent Behavior*, *58*: 101585. https://doi.org/10.1016/j.avb.2021.101585

when a design is powered enough to capture a 0.2 difference between treatment and control, it will also have the power to detect higher effects (e.g. 0.3, 0.7).

Power of covariates

Including covariates in the analytical strategy improves the precision of the estimations and, consequently, the power of the design. Incorporating additional participant or site characteristics as predictors makes it easier to detect the effect (if any). Therefore, the higher the power of covariates, the lower sample is often the outcome in the past, which is why the power of covariates tends to be higher when it includes a baseline measurement of the outcome. An advantage of the two outcomes measured above, and in general of soft outcomes, is that it is possible to estimate a baseline for them. However, there are other types of outcomes (usually hard outcomes) where the baseline doesn´t contribute to the power of covariates. This would be the case, for instance, when measuring the prevalence of domestic abuse. All mothers engaged with DART have been victims of domestic abuse, so this variable will show no variation at baseline, which does not add to power. Consequently, a trial measuring such an outcome would require a larger sample size.

While baseline data is normally collected through surveys, information about covariates, especially sociodemographics, can also be accessed through the administrative records of DART, although some capacity building would need to be included.

For the purpose of these power calculations, we will include two possible power of covariates rates: 0.3 for a scenario using “hard outcomes” and with few sociodemographic characteristics included, and 0.5 for a scenario with “soft outcomes” with baseline and several covariates. While we could potentially have a higher scenario with 0.7 of power of covariates, for instance, we prefer to be conservative in our estimations.

Intracluster correlation

For these power calculations, we will assume an intracluster correlation equal to zero. We have made this decision since the intervention is allocated at the mother level. While outcomes would be measured at both the mother and child level, most of the mothers participating also have a single child; therefore, there is no need to conduct a clustered randomisation, for instance.

Sample size

The sample of participants is especially important when designing the evaluation, both from an analytical and an operative perspective. Analytically, the sample size is the key factor that defines the effect size the design can capture. Operatively, it defines the length and cost of the data collection for the evaluation.

During the feasibility assessment, DART has been rolled out across 75 mothers and 75 children. This number covers all the NSPCC Hubs (including the one involved with co-delivering DART with a scale-up site), as well as the additional three scale-up sites. It is possible that there are more families served; however, additional sites did not respond to our data request and are therefore not included in the analysis.

The power of the design is defined based on complete cases, meaning the number of mothers who have completed the 10 DART sessions. We, therefore, must account for attrition around the participants' engagement. To approximate the rate of attrition of a full-scale trial, we will consider the intervention

take-up rate, as well as the rate of programme completeness. The take-up rate would be particularly important if the randomisation were to potentially happen after participants have been pre-assessed and assessed by the frontline workers. From data on the NSPCC's Hubs, we estimated a take-up rate of the intervention of 19%. This means that out of the 81 participants referred during the evaluation duration, only 15 of those mothers have received the DART intervention. The others are classified either as on "pre-engagement" or "assessment" or engaged in non-DART interventions at NSPCC. The scale-up sites only provided information on the participants who received the DART intervention, and we did not have the total number of participants referred or assessed prior to engaging with DART, for instance. Therefore, we based our estimations for take-up only on NSPCC's data. Regarding completeness rate, we found that, across all the sites, 84% of the mothers who start the 10 DART sessions complete them. This is promising since it means that the assessment period the case managers conduct makes it more likely that participants will complete the programme.

We have, therefore, included in our power calculations two scenarios of attrition rates. For a scenario where the randomisation happens immediately after referral, we consider an attrition rate of 65%, slightly higher than the take-up reported, to account for potential participants that could join DART after pre-engagement or assessment. For a scenario where the randomisation happens after the case worker's assessment, we will consider an attrition rate of 20%.

Power simulations

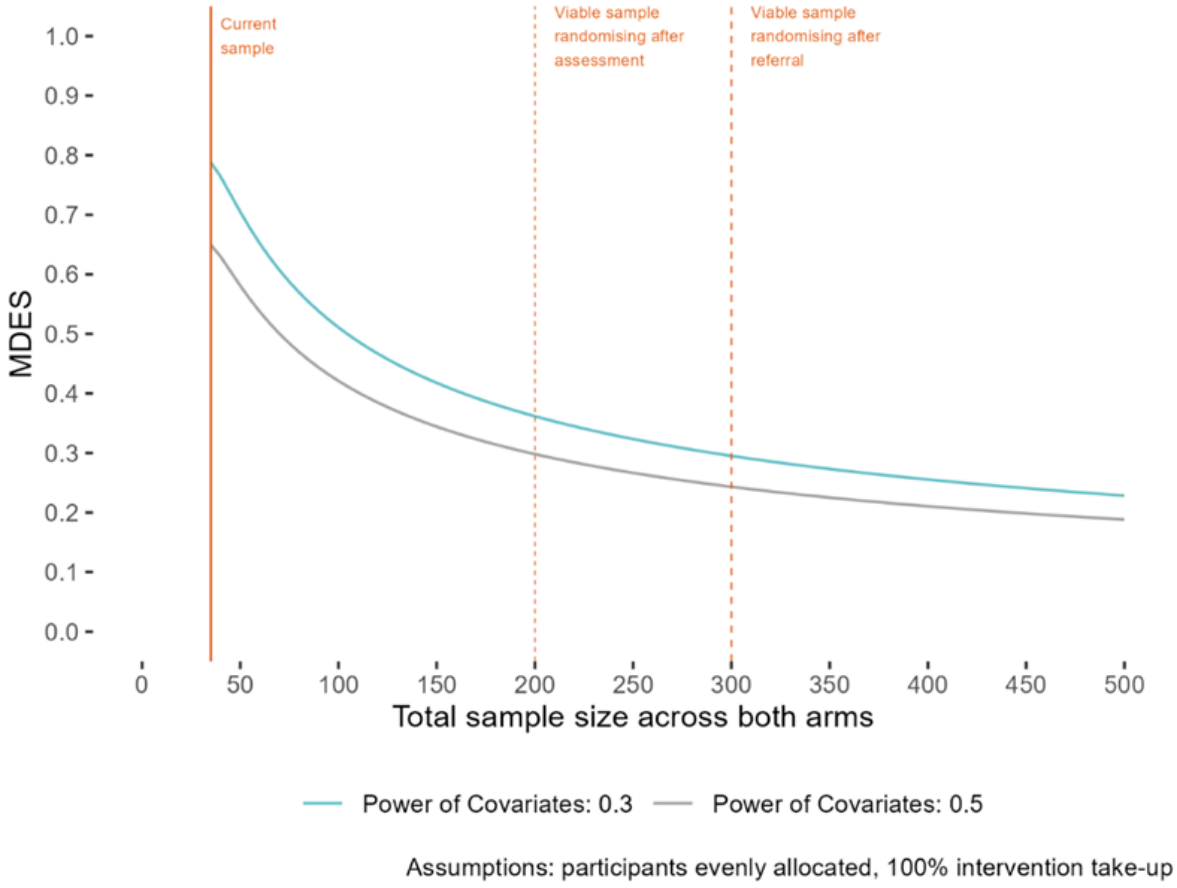
Table 11 summarises the power of a set of scenarios to evaluate the impact of DART. We have included what we call a feasibility flag. This is a marker of whether a defined set of conditions and sample sizes is likely to capture either a small-sized effect (0.2 Cohen's standard deviations) or a moderate-sized effect (0.4 Cohen's standard deviation). The flag would be red if capturing the effect size is unfeasible, yellow if there is a viable course of action which requires several changes and green if the design is well powered.

Considering the current numbers in the programme (75 families served), it will not be feasible to conduct a well-powered impact evaluation, as highlighted by the rows in red. Further scenarios are displayed to get an idea of the extent of the sample and the associated effect sizes. It is worth noting two simulated designs: if considering randomising after the referral process, we will require at least 300 participants in total to estimate an effect size of 0.4 or more. If randomisation were to occur after the assessment process, we would require at least 200 participants in total to estimate an effect size of 0.4 or more. From the literature review, we have seen that this type of intervention normally delivers smaller sample sizes than 0.4. Therefore, we will expect to work with higher sample sizes to ensure a robust design.

Table 11: Power simulation and feasibility

NUMBER OF PARTICIPANTS	ATTRITION	COMPLETE CASES	POWER OF COVARIATES	MINIMUM DETECTABLE SIZE EFFECT (MDES)	FEASIBILITY FLAG (assuming 0.2 size effect)	FEASIBILITY FLAG (assuming 0.4 size effect)
100	65%	35	0.3	**0.788**	🔴	🔴
100	20%	80	0.3	**0.569**	🔴	🔴
100	0%	100	0.3	**0.510**	🔴	🔴
100	65%	35	0.5	**0.649**	🔴	🔴
100	20%	80	0.5	**0.469**	🔴	🔴
100	0%	100	0.5	**0.420**	🔴	🟡
200	65%	70	0.3	**0.607**	🔴	🔴
200	20%	160	0.3	**0.404**	🔴	🟡
200	0%	200	0.3	**0.362**	🔴	🟢
200	65%	70	0.5	**0.500**	🔴	🔴
200	20%	160	0.5	**0.333**	🔴	🟢
200	0%	200	0.5	**0.298**	🔴	🟢
300	65%	105	0.3	**0.490**	🔴	🔴
300	20%	240	0.3	**0.330**	🔴	🟢
300	0%	300	0.3	**0.295**	🔴	🟢
300	65%	105	0.5	**0.411**	🔴	🟡
300	20%	240	0.5	**0.272**	🔴	🟢
300	0%	300	0.5	**0.243**	🟡	🟢
500	65%	175	0.3	**0.387**	🔴	🟢
500	20%	400	0.3	**0.255**	🔴	🟢
500	0%	500	0.3	**0.228**	🟡	🟢
500	65%	175	0.5	**0.319**	🔴	🟢
500	20%	400	0.5	**0.211**	🟡	🟢
500	0%	500	0.5	**0.188**	🟢	🟢

Figure 7: Impact evaluation feasibility – power simulations curve



Progression criteria: sample size for the proposed design

This criterion considers if the currently available sample size is adequate for an experimental or a QED design and whether there are clear routes to create an adequate sample size. As discussed above, the current numbers of families accessing DART would not be sufficient to enable us to conduct a well-powered impact evaluation (it is also important to note that these numbers are the total families accessing the service over **roughly 20 months,** which is longer than a traditional RCT or QED design would normally recruit for, suggesting an even greater increase may be needed). Given the ad-hoc and inconsistent nature of referrals at the case study sites, there are not currently clear routes to create an adequate sample size.

Research question 6: to what extent would experimental or quasi-experimental methodologies be acceptable to referrers, NSPCC and DART staff?

Acceptability of a randomised control trial or a quasi-experimental design to staff

Acceptability of an experimental or quasi-experimental methodology for evaluating DART varies among DART staff and referrers.

Participants generally recognised the benefits of evaluations and research, including that they can help increase funding and referrals, improve service quality and further improve the reputation of DART by proving its impact.

"I think it's always good to have research to back stuff up, isn't it? It makes you have a little bit more confidence when delivering it if you can say, 'Actually, we know this works, and this is why we know it works.'" (DART Practitioner)

Participants shared a desire to gain deeper insights into the long-term impacts of DART and its effectiveness within different communities.

However, participants also raised practical concerns and challenges associated with researching DART. They emphasised the need for a trauma-informed approach due to the sensitive nature of the topic. Additionally, they emphasised the need for practitioners to buy in to the research process, which may require input from senior staff.

A major concern of many practitioners was resource constraints; they expressed the need to secure additional funding for both the sustainability of DART provision and to enable practitioners to adequately engage with the research.

Additionally, establishing strategic commitment to the longevity of DART provision within NSPCC was considered important before taking research forward.

However, it is worth noting that there is also a view among participants that they already know that DART works. This view stems from two perspectives: frontline staff, who rely on their direct experience to validate DART's effectiveness, and senior staff, who question the need for a new evaluation, believing it would not add new insights beyond what previous evaluations of DART have already shown.

"Is it needed? What is going to be different that you're going to evaluate to what we already have evaluated, I guess". (NSPCC Senior Staff Member)

"We know from the mums that we work with how impactful DART is; like, they tell us, we see them…I don't need a study to tell me, not in a nasty way…but like I, you know, we see them, we see the positive impact". (Practitioner)

This view is countered in the section on "Evidence and rationale for DART", which highlights the limits of the current evaluations of DART and what an impact evaluation could add.

The perspectives on RCTs were notably more negative than the views expressed about research in general. While no participant described themselves as opposed to research in principle, very strong negative views were expressed about the RCT approach described.

A key concern among participants, across sites and irrespective of seniority, was the ethics of conducting an RCT with DART clients (mothers and children). Some felt that this approach could be exploitative towards clients.

"So, I just feel as though we're using their unfortunate circumstances to get results, and to say to them, 'In the future, it will help other people, but it's not going to help you.'" (DART Practitioner)

Others raised ethical concerns that some families would be denied access to or left waiting for a service that they needed. DART was seen as a "needs-led" service, which is difficult to combine with randomisation.

"Being totally honest, I don't like the idea of that because I don't like the idea of people being on waiting lists. I think it's really important that they get the right support when they kind of need it". (DART Practitioner)

Building on this, some of the ethical concerns expressed by participants were seen as fundamentally contradicting their understanding of their role as practitioners or NSPCC staff. As one senior staff member put it:

"Through our own personal passion and belief in the organisation about who should receive DART, I just don't think there'd ever be an appetite; it feels wrong…from a moral perspective". (Senior NSPCC Staff)

Previous experiences with other evaluations also tainted their view of RCTs. At one site, a practitioner described that they *"haven't heard anybody have a positive word to say about* [an RCT]*"*, while a senior staff member expressed concern about whether DART would get support from the leadership or would even be strategically useful.

For some, the level of negativity they felt towards RCTs was such that they expressed themselves using inflammatory language, drawing parallels between the methodological approach and historical atrocities. While extreme, this does help illustrate both the depth of feeling and the visceral nature of the response that participants had across locations, site types and seniority levels.

It should be highlighted that the negative response to RCTs was not universal. There were positive views about using an RCT methodology, such as its ability to demonstrate the impact of DART and lend it extra credibility. Others saw RCTs as a plausible methodology, even if they did not actively endorse it in this instance.

In exploring the possibility of carrying out an RCT, we also sought input from staff and referrers regarding the use of a waitlist design. While some of the sites had some form of waitlist in place for families waiting to access DART, these waitlists were generally described as relatively short. When asked about what length of waitlist would be appropriate if it were used as part of an RCT, there was no consensus among the participants. However, a waitlist period of one year was broadly seen as too lengthy, and delays in accessing DART that were considered acceptable ranged from three to six months.

Some participants were open to the idea of using a waitlist as part of an RCT design, particularly when the waitlist already existed due to factors outside of the research. Participants also acknowledged that a delay might not be problematic for some service users or clients. However, they expressed concerns about the potential negative effects of a waitlist on families. A primary concern was the time-sensitive nature of the support that families affected by domestic abuse need:

"We're dealing with domestic abuse. It can turn on a sixpence. I would not want anybody sat on a waiting list". (Practitioner)

Participants felt that families left without support might experience changes in their situation, leading to their disengagement or making them unsuitable for DART in the future, for instance, if the situation deteriorated and their children were taken away. Another key concern was that families could be at additional risk while waiting, such as their mental health deteriorating or the mother reconciling with her partner, further risking their safety.

"They might go back to their partner because they've not had that support, so then they're putting the children and themselves at risk again, and then we couldn't work with them at all because they'd be back together in the same household as the alleged perpetrator". (Practitioner)

Some participants suggested they might be more open to a waitlist approach if they had the ability to choose who goes on the waitlist or if there was some flexibility in its implementation. However, this would not be possible within the current method.

Participants were also concerned that requiring families to wait could damage the relationship between the families and DART staff or might result in fewer referrals from referrers to the programme.

One solution that participants proposed that they felt could mitigate concerns about a waitlist approach would be to provide families with support while they are on the waitlist. This support could include regular check-ins with families to manage risk or the provision of other services that families are likely to receive anyway while waiting for DART.

Additionally, participants also raised practical challenges that might make it difficult to use an RCT approach to evaluate DART. A key practical challenge was the need to recruit a sufficient number of participants to meet the requirements of an RCT. Some participants expressed difficulty in already obtaining enough families for a DART group, and the need for even more participants was seen as a significant concern.

"[We] struggle now to get enough families for a group, and actually we would need double that number, and that's my biggest concern…that it's just not achievable". (Senior NSPCC Staff)

Additionally, the requirement to run DART within an appropriate age group further complicated the task of getting the numbers needed. As per the NSPCC manual, DART groups are generally split between older and younger children (see the description in the Introduction section of this report).

Staff also described the need to ensure that referrers were bought into the process, and they expressed concerns that referrers might be less likely to refer families to DART if an RCT approach was present.

Concerns about randomised control trials

The section above outlines the views of practitioners, senior staff and referrers of RCTs. Their views include some perceptions of RCTs that would be useful to address here.

A large number of the concerns raised about the ethics of conducting an RCT relate to the view among staff that DART is better than business as usual. If this is the case, then denying access to a service that improves outcomes for families is considered unethical. This is a perspective not uncommon from those delivering an intervention. However, as discussed in the section on "Evidence and rationale for DART", the current evidence does not conclusively establish that the programme is better than business as usual. Given this, assigning families to the control group would not be denying access to an effective service but would rather be denying access to a service that may or may not be better than business as usual.

It is also worth noting that an RCT design would not require that families assigned to the control group do not have access to support. For most RCT designs, those in the treatment group would be able to access business as usual support, i.e. the support they would receive if DART was not running.

As discussed below in the recommendation section, addressing these points when building buy-in for an RCT approach would resolve some of the acceptability concerns that have been raised specifically in relation to RCTs.

In contrast, QEDs were, on the whole, viewed more positively than RCTs, although participants raised practical considerations and concerns about how valid a comparison group would be.

A QED approach was described as preferable to an RCT, particularly because it allowed all families to receive support, and this was viewed as ethically more comfortable for staff and referrers.

"This feels more comfortable in the fact that it's not excluding people who may have been referred to DART". (Senior NSPCC Staff).

Participants also noted the potential for QEDs to provide long-term findings on the impact of DART,[9] which might also be able to unlock additional funding. As mentioned above, this long-term outcome from DART was a key area of interest for any future research.

Negative views of QEDs focused on methodological concerns, as some participants questioned whether it would be worth doing, considering that, methodologically, it was considered less robust than an RCT. There were also concerns about the time and resources required for conducting a QED.

Practical concerns were also a consideration for participants when discussing a QED approach. Participants were aware of the difficulty of finding a suitable comparator group for DART participants. This was complicated by the fact that individuals with experience of domestic abuse often protect their privacy to keep themselves safe, making it difficult to identify and track them over time. While they acknowledged that some organisations might have access to families that could be in a comparator group, the number required and the level of information sharing needed to create a comparator group presented significant challenges. Additionally, participants also noted that the competition between organisations providing services might mean that conducting a QED – which might prove DART to be better than the "business as usual" they provide – would result in an organisation not wanting to be involved.

Another concern was around how comparable a comparison group would be. Things that would make families ineligible for DART might also endanger their ability to be a comparator for DART service users/clients, such as being in an ongoing relationship with the perpetrator or going through court proceedings.

"[I] think if a lot of them that aren't eligible, it's for reasons like that; they are going to probably have very different results anyway compared to a group that are further along in their journey, do you know what I mean?" (Practitioner)

Progression criteria: acceptability of the design

Given the strength of feeling expressed about an RCT approach, an RCT design is not currently acceptable. In contrast, a QED design could broadly be considered acceptable to staff and referrers. Concerns about

[9] While RCTs can also provide this insight, due to the controversial nature of RCT design in this setting, the design we suggested to participants was mainly a waitlist design, which would not allow long-term outcomes to be tracked.

the QED approach were largely practical in nature and didn't significantly affect the acceptability of the design.

Relationship between evaluators and the National Society for the Prevention of Cruelty to Children

For an impact evaluation to be possible, there needs to be a constructive and open working relationship between the evaluator, the NSPCC and scale-up sites. During the feasibility study, the relationship with the NSPCC and scale-up sites was generally supportive, but resource constraints have limited their ability to appropriately support the study. Addressing capacity issues and enabling staff to assign adequate time and resources to supporting a full experimental or QED design is important. Additionally, sufficient resources need to be available for NSPCC and scale-up sites to drive future evaluation work forward.

To consider this, we provide a summary below of our relationship with the NSPCC and scale-up sites during the feasibility evaluation, along with an assessment of the buy-in and available resources for a future project.

Throughout the feasibility study, the evaluation team collaborated with members of the central NSPCC team, NSPCC Hub staff and four scale-up sites. The organisations have been consistently helpful and supportive of our work. However, it has been clear that there is pressure on resources for both the scale-up sites and the NSPCC, and in some instances, this has limited their ability to support the study in a timely manner. This limitation became particularly pronounced when accessing administrative data, as it required staff to carry out tasks that were out of the scope of their usual work. We also asked all scale-up sites (including those not taking part as a case study for this research) to complete a very short online questionnaire to let us know what data they collected, but only one site was able to do this.

Resource constraints have significantly affected participants' perspectives on the desirability of a further evaluation of DART. Their support for further research was contingent on the availability of resources and whether it would put extra burdens on their team:

"…as long as there is not going to be any harm to the families involved or too much further stress to practitioners, then I can't see any issues with evaluations". (Practitioner)

For scale-up sites, securing additional funding was suggested as a potential solution to resolve the resource issue, as funding is often the major concern in their work.

Throughout the feasibility study and interviews, it became evident that there were varying levels of enthusiasm for a further evaluation of DART. This stemmed from either the belief that the evaluation would not contribute anything further to their understanding, given previous evaluations, or strong ethical and/or practical concerns about the proposed methods. These views, combined with resource constraints, have made staff less eager and more anxious about committing the time and resources to make an evaluation work.

"We'll have to weigh up, like, all the things that we're working on at the moment, and, like, do we really want to put the energy into this or into something else. It's really hard to say". (Senior NSPCC Staff)

Progression criteria: working relationship between the evaluator and the National Society for the Prevention of Cruelty to Children

Given the above, we would suggest the progression criterion concerning the working relationship between the evaluator and the NSPCC has not been met. There are varying levels of commitment and enthusiasm for conducting a further evaluation of DART, and significant resource and capacity constraints would make it challenging for NSPCC Hubs and scale-up sites to effectively support such an evaluation.

Conclusion

Figure 8: Summary of feasibility study findings

Research question	Finding
1. How is the DART programme currently being delivered at different sites?	DART is delivered at NSPCC Hubs across the country, as well as 25 external scale-up sites, by trained DART practitioners. External sites are provided with a programme manual, a day and a half of in-person training for DART practitioners, an information pack with a guide to implementing the service and ongoing telephone support. The DART sites show some variation in delivery settings, with some operating either on-site or in collaboration with local domestic abuse organisations. While most sites offer the programme two to three times per year, some have faced challenges in securing referrals. This has led to significant adaptations of the DART model, with some hubs serving as few as one or two families. Although five of the seven sites included in the study are in regions with domestic abuse rates higher than the national average in England, the demand for DART (i.e. those attending the DART sessions) did not reflect these statistics. This discrepancy was due to various factors, including the presence of other established domestic abuse services, complex eligibility and suitability criteria that further limited the number of families accessing DART, the competitive funding landscape in the domestic abuse sector, which prevented partnership with other established domestic abuse services, and the funding cuts that led to the closures of key referral partners further limiting the number of referrals. All sites adhered to the eligibility criteria set by the NSPCC and outlined in the manual (this includes eligible children to be aged 7–14, lived with domestic abuse experience, assessed as harmed by this, perpetrator no longer part of household – see full list in Table 1). However, practitioners used additional factors to assess "readiness" for DART post-referral, which were based on their professional judgement. Professional Judgement created ambiguity in the criteria within and across sites. Inadequate referrals were another barrier to consistent programme delivery, largely due to significant funding constraints in the domestic abuse sector. The use of the DART manual varied among sites. Most practitioners found it a valuable resource, but some found it outdated and complex at times. Sites adapted the manual to better cater to the unique needs of families, including addressing language and disability barriers. Finally, staffing and financial constraints caused significant challenges, as the programme required four practitioners per DART group. Many practitioners were understaffed, managing DART alongside up to four other responsibilities and significant caseloads. Progression criteria: the research findings identified areas for improvement in meeting certain progression criteria, mainly around establishing clear referral routes and improving guidance on eligibility criteria and suitability criteria. While the study recognises adherence to NSPCC eligibility criteria across all sites, it highlights challenges around suitability criteria and variations in professional judgement across sites. Recommendation: there should be a balance between allowing practitioners to exercise professional judgement and also providing clearer guidance on the interpretation of DART eligibility and suitability criteria. Additionally, funding uncertainties are a fundamental constraint in the domestic abuse sector; funders should consider increasing funding and exploring partnership work for DART to reduce the impact of the competitive funding landscape. This would enable more suitable referrals to DART and reduce the number of families that are unable to progress into DART as a result of the assessment process and suitability criteria. More details on the recommendations are provided in the conclusion section.
2. What is the user experience of DART?	The study did not directly engage with mothers and children as it was not part of the scope. However, it analysed nearly two years of DART programme data, during which it served a total of 75 families, each comprising a mother and a child who had experienced domestic abuse. These families were served across all four NSPCC Hubs (including the one involved with co-delivering DART with a scale-up site) and the additional three scale-up sites. The average age of the mothers in the sample was 41 years, ranging from 29 to 50 years, while the children's average age was 11, with the youngest being seven and the oldest 14. Among the 75 families analysed, where information on the referrals for 21 families was unavailable, the most common referral was through local authorities, accounting for 36% of the families, either

	through the Integrated Front Door, children's social care services or early help. Self-referral was the second most common pathway. In terms of completion rates, the majority of families that started the DART programme tended to complete it. On average, families completed 9.5 out of the 10 sessions, with only 16% not completing the full 10 sessions during the evaluation period. Qualitative interviews with DART practitioners also supported these findings, indicating low attrition rates and high programme completion. Overall, there are minimal differences between White participants and those from an ethnic minority background, with at least 80% of the families completing the programme. Perceived impact of DART on families: measuring outcomes and engaging directly with mothers and children fell outside the study's scope. However, the study did seek input from practitioners and referrers. According to both, the programme consistently delivered significant benefits to mothers and their children. The benefits observed included the improvement of the mother–child relationship and better understanding, communication and connection within families. DART practitioners perceived that the programme also improved emotional management and built resilience, self-confidence and self-esteem in participants, and mothers seemed to gain a better understanding of available support services in their local area. Recommendation: practitioners consistently shared that the programme delivered significant benefits to the families participating in DART, which is encouraging. The study found variations in the way outcomes were measured across sites, which makes it difficult to fully assess the programme's impact. Standardising outcome measures across sites and involving families in future research (e.g. pilot evaluations) will provide better insights into the programme's impact.
3. What are the barriers to DART's cohort reflecting the ethnic diversity of the communities it works in? What strategies could be used to address these barriers?	The study found that most sites included in the study are in predominantly White (British) areas. Only one area included in the study had a White population of less than 75%, according to the 2021 census. Of the 75 families served by DART, around 85% self-identified as White, and 15% of the sample were families that self-identified as non-White, including Black Caribbean, Black British and British Asian. Interviews with DART practitioners and referrers revealed that DART participants are primarily White British, and English is their first language, which aligns with local demographics. However, this data is at the local authority level, and diversity within local authorities – from one neighbourhood to the next – can vary considerably, making it challenging to gauge how well DART cohorts represent their local communities. To illustrate, some sites reported the presence of growing minority communities that were not adequately captured in the census data as the local landscape changes. Given this, DART practitioners and referrers identified several barriers preventing the participation of families from diverse backgrounds, even in areas where ethnic diversity is more prevalent, including cultural differences impacting the understanding and willingness to address domestic abuse, low awareness of DART and language. Some sites have started engaging with local minority communities to address these barriers, such as organising community-specific events and providing information in various languages. Translators and interpreters have been employed to ensure non-English speakers or those requiring sign language interpretation can participate. However, some DART sites shared that integrating translators and interpreters into groups can be challenging, and the costs of translating materials can also be an additional barrier for sites with resource constraints. The lack of diversity among DART practitioners, who are predominantly White, is also considered a potential barrier to building trust and rapport with families from diverse backgrounds. Some practitioners suggest expanding the eligibility criteria to include same-sex couples, male survivors and non-biological mothers to enhance diversity. However, concerns exist about the demand from these groups and the appropriateness of DART as a support option. To better reach non-English-speaking participants, running specific DART sessions for non-English-speaking participants who make up a large proportion of the local community has been suggested, although some sites are not able to do this themselves due to resource constraints. Progression criteria: although the majority of DART participants and practitioners are White British, aligning with local demographics, practitioners have identified the presence of minority communities and some barriers preventing the participation of families from diverse backgrounds. Recommendation: partnering with well-established organisations with ties to diverse groups, such as minority-led domestic abuse organisations or specialist domestic abuse services, is recommended to improve outreach (see the "Recommendations" section for more detail).
4. To what extent is an experimental or quasi-experimental methodology for an impact evaluation of the DART programme practically possible?	This study considered several factors when assessing the feasibility of implementing an experimental or quasi-experimental methodology for assessing the impact of DART, including referral routes, eligibility criteria, barriers to randomisation, constructing a comparison group, sample size and measurable outcomes. Referral routes: a number of DART sites rely on limited and ad-hoc referrals, which makes it challenging to gather a sample size suitable for an RCT or a QED due to insufficient numbers. Inconsistent and limited referrals to DART are influenced by funding cuts in the domestic abuse support services. Developing reliable referral pathways is key to increasing the potential sample size.

	Eligibility: eligibility criteria for DART are well defined by the NSPCC. However, practitioners also rely on additional criteria, termed "suitability criteria" in this study, to assess the "readiness" of families to access DART after they are referred. These suitability criteria vary across sites as they are based on professional judgment to ensure that assessments are person-centred and take into account individual circumstances beyond the official eligibility criteria. While the use of professional judgement and suitability criteria is important in the domestic abuse sector, this may have limited the number of families accessing DART. Barriers to randomisation: the study found that randomisation could be added to the DART process at two points: at the referral point or after the pre-DART assessment phase. There's no clear consensus among practitioners, staff and referrers, but for analytical reasons and to maximise power, the study recommends that randomisation should happen after the assessment processes. Constructing a comparator group: to construct a comparator group for a QED, several approaches were considered, including DID, matched comparator group, matched DID and regression discontinuity designs**.** The study found that using a matched comparator and/or matched DID approach would be methodologically possible for the DART programme. However, this will require resources and relationship building with organisations that would have access to potential comparator participants. Given the under-resourced nature of the sector, funding would likely be required to support organisations to participate in the evaluation. Data management: the study found that the level of data management is relatively low across a number of sites due to capacity constraints. Outcome measures: the study found that sites vary in the outcomes they measure, and data quality across sites is inconsistent. NSPCC consistently uses validated scales to measure two outcomes (self-esteem measured through the Rosenberg Self-Esteem Scale and attitudes and behaviours measured through the SDQ). In contrast, the other sites show a lack of standardisation, relying on a self-developed questionnaire and free-text feedback. Although there is no current measure of parent–child relationship, in the last DART internal evaluation, the Parent-Child Relationship Inventory was used. This was discontinued after the evaluation ended. Measuring the quality of the mother–child relationship is essential for impact assessment. Progression criteria: the feasibility of conducting an impact evaluation depends on several factors, including sample size, randomisation approach, comparator group construction and data collection methods. Significant resources are needed to overcome challenges in terms of sample size and better management of administrative data and outcome measures. While an RCT might be more manageable than a QED approach, it faces challenges related to acceptability among staff and referrers. At this time, the study found that an RCT or a QED would only be possible if there is a commitment to overcoming existing constraints and challenges. Successful implementation of an impact evaluation would demand significant resources, particularly with regard to increasing the potential sample size (to at least three times the number DART currently serves). This would entail DART sites either increasing their caseloads or new sites delivering DART. However, the resource-intensive nature of DART, coupled with the time needed for site establishment and referral route development, suggests that a significant time investment is required before DART services can support a feasible impact evaluation. Recommendation: in order to make an RCT or a QED practically possible, referral routes would need to be developed. Building partnerships with the wider domestic abuse sector, with the support of NSPCC, could be one way of achieving this goal. In order to more accurately gauge the potential sample size available for an RCT, a clearer understanding of "suitability" among DART providers would need to be developed. Additionally, validated scales for mother–child relationship assessment need to be identified to standardise outcome measures across all sites. If an RCT approach is used, the randomisation point needs to be agreed upon; this study recommends that randomisation should happen after the assessment processes, and if a QED approach is used, significant time and resources need to be committed to construct a comparator group. Finally, capacity building (including in data management) to improve data quality is recommended to ensure a successful evaluation (see the "Recommendations" section for more detail).
5. What sample size would the DART programme currently be able to provide for an impact evaluation, and are there credible routes to increase the potential sample size? What is the estimated sample size required to achieve adequate statistical power for a future impact evaluation of the DART programme?	The power calculations presented in this study were informed by various parameters, including MDES, as reported in the literature for similar programmes, estimated power of covariates, intracluster correlations when relevant and attrition rates. Attrition rates were determined based on two scenarios: randomisation after referral (65% attrition) and randomisation after assessment (20% attrition), taking into account the response rates reported in the administrative data analysis. Additionally, power calculations were conducted across multiple scenarios to reflect the requirements for evaluating DART in a full-scale trial. Considering the current status of the programme, this study found that the current number of families accessing DART would not be sufficient to conduct a well-powered impact evaluation. To conduct a robust impact evaluation of DART, DART would need to serve at least three times the number of families it currently does to make the statistical design viable. Progression criteria – sample size for the proposed design: this criterion considers if the currently available sample size is adequate for an experimental or QED design and whether there are clear routes to create an adequate sample size. The current number of families accessing DART would not be sufficient to enable us to conduct a well-powered impact evaluation. Given the ad-hoc and

	inconsistent nature of referrals at the case study sites, there are not currently clear routes to create an adequate sample size. Recommendation: to conduct a robust impact evaluation of DART, a substantial increase in the number of families served by DART would be required. DART would need to serve at least three times the number of families it currently serves to make the statistical design viable (see the "Recommendations" section for more detail).
6. To what extent would experimental or quasi-experimental methodologies be acceptable to referrers, NSPCC and DART staff?	Acceptability of an experimental or quasi-experimental methodology for evaluating DART varies among DART staff and referrers. Participants generally recognised the benefits of evaluations and research, including that they can help increase funding and referrals, improve service quality and further improve the reputation of DART by proving its impact. A major concern of many practitioners was resource constraints, which need to be addressed; there was an apparent requirement to secure additional funding for both the sustainability of DART provision and enabling practitioners to adequately engage with the research. Views on methodology: participants' views on further evaluation were influenced by resource availability and ethical concerns. The perspectives on RCTs were notably more negative (though not universally held), mainly due to ethical concerns around denying or delaying access to the programme for families allocated in the control group but also practical concerns such as the need to recruit a sufficient number of participants to meet the requirements of an RCT. In contrast, a QED design was considered more acceptable, primarily because it allowed all families to receive support. A constructive and open working relationship is crucial for an impact evaluation to be successful. During the feasibility study, the relationship with the NSPCC and scale-up sites was generally supportive, but resource constraints have limited their ability to support the study. Progression criteria: given the strength of feeling expressed about an RCT approach, an RCT design does not currently seem acceptable, while QEDs are more acceptable. Recommendation: for an RCT to be acceptable, efforts should be focused on building buy-in for an RCT design at all levels of the organisations, given that negative views about RCTs were widespread. Addressing key concerns involves conducting an RCT in an ethical manner that aligns with the ethos of DART practitioners. Emphasising the long-term benefits for families resulting from the evaluation could also garner support for the RCT approach. To support staff and stakeholder participation in an evaluation, the first step would be providing the NSPCC and scale-up sites with significant extra resources (see the "Recommendations" section for more detail).

Key recommendations

In order for an impact evaluation of the DART programme to be feasible, we recommend the following steps. These are covered below under the research question they relate to.

Research question 1: how is the DART programme currently being delivered at different sites?

The delivery of the DART programme varies across sites, with some sites having more established referral pathways and more consistent practices than others. The funding landscape for domestic abuse services is significantly limited and competitive, which creates challenges for DART referrals. Additionally, DART has a range of eligibility and suitability criteria that are not always applied consistently across sites. DART practitioners rely on their professional judgement – an essential part of the domestic violence sector – to assess whether families are ready for the programme, which can lead to differences in decision making across sites. Finally, the demand for DART does not always match the need, with some sites receiving fewer referrals than expected.

Recommendation: establish clearer eligibility criteria and referral routes.

There should be a balance between allowing practitioners to exercise professional judgement and also providing clearer guidance on the interpretation of DART eligibility and suitability criteria, including in relation to mothers requiring previous domestic abuse work to access DART. In order to make a full evaluation of DART possible, clear and stable referral routes should also be established. One way of securing this is for DART sites to develop stronger relationships with referral partners, including the wider domestic abuse sector, to improve the number of referrals.

Recommendation: increase funding for the sector and partnership work.

Additionally, funding uncertainties are fundamental constraints in the domestic abuse sector; funders should consider exploring partnership work for DART to reduce the impact of the competitive funding landscape. This would enable more suitable referrals to DART and reduce the number of families that are unable to progress into DART as a result of the assessment process and suitability criteria. Finally, DART sites should develop networking and communication strategies with other domestic abuse services to increase partnership work and avoid duplication of work.

Research question 2: what is the user experience of the DART programme?

Recommendation: standardise outcome measures across the sites.

As previously mentioned, directly engaging with families was out of the scope of this study. However, practitioners consistently shared that the programme delivered significant benefits to the families participating in DART, which is encouraging. The study found variations in the way outcomes were measured across sites, which makes it difficult to fully assess the programme's impact. Standardising outcome measures across sites and involving families in future research (e.g. pilot evaluations) will provide better insights into the programme's impact.

Research question 3: what are the barriers to DART's cohort reflecting the ethnic diversity of the communities it works in? What strategies could be used to address these barriers?

Recommendation: improve outreach by partnering with established and diverse minority-led organisations to better reach diverse communities and recruit DART practitioners from various ethnic backgrounds.

Although the majority of DART participants are primarily White British, and English is their first language, which aligns with local demographics, some sites have reported the presence of growing minority communities not adequately captured in the census. Some DART sites are already using various approaches to reach out to diverse communities from ethnic minorities. Partnering with well-established organisations with ties to diverse groups, such as minority-led domestic abuse organisations or specialist domestic abuse services, and recruiting DART practitioners from various ethnic backgrounds is recommended to improve outreach.

Research question 4: to what extent is an experimental or quasi-experimental methodology for an impact evaluation of DART possible?

Recommendation: develop referral routes to increase the number of referrals.

In order to make an RCT or a QED practically possible, referral routes would need to be developed. Some sites reported on their efforts to inform potential referrers of DART, going to local authority teams and police forces to present the programme. Other hubs described attempted partnership relationships with local authority teams and national domestic abuse organisations. These have been achieved to varying degrees of success.

Funding, however, appears to be a significant barrier to such relationships being strengthened for the site running DART or for the organisation unable to enter into partnership for fear of losing its own funding. While the competitive funding landscape itself might not be overcome, support to build and negotiate referral pathways could be facilitated by the NSPCC.

It may be possible, for example, for the NSPCC centrally to begin discussions with national-level organisations or those supporting minority communities and to support DART sites and partner organisations alike to institutionalise referral pathways suitable to their local landscape. While local domestic abuse services vary from site to site, the NSPCC could, through its already established networks of DART practitioners, learn from existing experience to provide guides and hands-on support to negotiate the establishment of pathways and the sharing of referrals.

Recommendation: eligibility, suitability and the assessment process could be clarified.

In order to more accurately gauge the potential sample size available for an RCT, a clearer understanding of “suitability” among DART providers would need to be developed. The existing network of DART practitioners could be drawn upon to define suitability criteria used across sites, of which there are multiple, and where and how less commonly used criteria could be brought into a shared understanding of suitability. Drawing on this network could also provide an understanding of how much “suitability” limits demand for DART.

Suitability depended, for many practitioners, on the “readiness” of a family; it needed to be the “right time” for a family to access DART. This depended on various factors, including housing stability and police or court involvement. However, DART’s timing as a recovery service creates additional complexities and significantly shapes how many families of those reporting domestic abuse incidents could be suitable for

DART; some may never want or be ready to access a domestic abuse recovery service. It may be worthwhile to consider what constitutes the "right time" for families to access DART. While some flexibility would be needed given the individual judgments made by practitioners and additional support put in for some families, doing so could provide a more definite idea of the size of the population that *could* benefit from DART.

The assessment process was described as time-consuming, especially at NSPCC Hubs, where NSPCC *and* DART assessments are carried out – both of which include at least three stages. Practitioners reported duplicating write-ups, and senior staff described how DART assessments were expected to last for eight hours. For practitioners delivering other services and working with significant caseloads, this is one factor limiting the possibility of increasing DART provision.

Scale-up sites are not required to conduct eight hours of assessments for DART; instead, they are expected to assess the safety and readiness of families. To ease pressure on hub staff and, in turn, potentially increase DART provision and the sample size available for an RCT or a QED, it would be valuable to investigate whether and how assessment processes at NSPCC Hubs, in particular, could be simplified.

Recommendation: if an RCT approach is used, the randomisation point needs to be agreed upon.

A considerable amount of time and resources are put into the assessment process for DART, as described above, and this process itself may have an impact on participants. If an RCT is going to be carried out, a decision should be made about whether the assessment process is considered to be part of the DART process, and therefore, baseline and randomisation need to happen before the assessment takes place. Work would also need to be done to address the challenges posed depending on what randomisation point was chosen. If the referral point is chosen at the randomisation point, this would lead to greater sample sizes being needed.

Recommendation: if a QED approach is used, significant time and resources need to be committed to construct a comparator group.

If a QED approach were taken to evaluate DART, the feasible approaches to constructing a comparator group would be a matched comparator or matched DID approach. However, both of these approaches would require a large comparator group to match from, which would be difficult to construct. In order for this approach to be possible, significant work would need to be carried out to build relationships with national and local organisations that provide services to, or interact with, families with experiences of domestic abuse. This would require a significant commitment of time and resources.

Recommendation: appropriate outcome measures need to be identified and agreed upon.

To conduct a full-scale RCT, it will be necessary to conduct a rapid literature review to identify highly reliable and validated scales to measure the quality of the mother–child relationship. It might also be useful to conduct a process of cognitive testing over the identified scales to ensure they are appropriate for working with survivors of domestic abuse, as such scales could not be validated in vulnerable populations. Given the differences in data management and outcome measurement across sites, it may be necessary for the evaluator not only to design the survey but to collect the data as well. It's also important

to encourage the three sites (excluding NSPCC) to start collecting the Rosenberg Self-Esteem Scale and the SDQ as part of their normal practices.[10]

Additionally, the overall level of data management is relatively low. During the descriptive data analysis in RQ**Error! Reference source not found.**, several concerns regarding the quality of the information provided were raised. It is also worth noting that the tasks involved in extracting information to inform this feasibility assessment appeared to represent a significant burden for the delivery organisations. Therefore, if we anticipate that a full trial would involve incorporating administrative data from the sites or integrating data collection as part of their regular engagement with the participants, it would be necessary to invest in capacity building and allocate additional resources to support these efforts.

Research question 5: what sample size would the DART programme currently be able to provide for an impact evaluation?

Recommendation: the potential sample size needs to be substantially increased.

To conduct a robust impact evaluation of DART, a substantial increase in the number of families served by DART would be required. The required increase is significant, as DART would need to serve at least three times the number of families it currently serves to make the statistical design viable, assuming all who took part in DART are involved in a subsequent evaluation.

Research question 6: to what extent would experimental or quasi-experimental methodologies be acceptable to referrers, NSPCC and DART staff?

Recommendation: for an RCT to be acceptable to staff and stakeholders, building buy-in needs to be a priority.

Significant changes would be needed to reach a point where an RCT design would be acceptable to staff and stakeholders.

Efforts should be focused on building buy-in for an RCT design at all levels of the organisations, given that negative views about RCTs were widespread. The strength of participants' views and the belief among practitioners that the RCT approach contradicts the ethos of their work mean that building buy-in will likely take some time. Staff and stakeholders/referrers will need to be convinced about the value of an RCT. Engaging colleagues within the NSPCC and scale-up sites is proposed as a strategy to build goodwill among staff.

Key points to address included conducting an RCT in a manner that doesn't create unsolvable ethical concerns or contravene the ethos of DART practitioners. Possible ways to achieve this could include using a design that leveraged existing waitlists. However, this would require a significant increase in appropriate referrals, as currently, there are not sufficient referrals for a waitlist, and referral channels are ad hoc and

[10] The Rosenberg Self-Esteem Scale and the SDQ are two widely used validated scales to measure the impact of different type of programmes. If an impact evaluation is to be conducted, we will require all providers of DART to measure outcomes in the same way. Using these two scales, as NSPCC already does, will provide good-quality information that can even potentially reduce costs for data collection. It will also facilitate comparing outcomes across providers as well as following the trajectory of the overall programme and participants across time. Furthermore, the exercise of embedding these scales in the sites can potentially improve the data management and data structure within each of the sites.

limited (see the section above on referral routes). Emphasising the long-term benefits for families that would come from the evaluation could also help build support for the RCT approach.

Recommendation: staff and stakeholder participation in an evaluation would be facilitated by extra resources and an understanding of the value additional evaluations would add.

To support staff and stakeholder participation in an evaluation, the first step would be providing the NSPCC and scale-up sites with significant extra resources. This would allow them to prioritise supporting the evaluation.

In addition to this, efforts should be made to work closely with the NSPCC and scale-up site staff to develop an understanding of what additional value an RCT or a QED evaluation could bring beyond what has been achieved in their previous evaluations. This would require some adjustments to the staff's current understanding of the strength of evidence regarding DART. While there is a strong belief in the effectiveness of DART based on existing evidence, from an evaluation perspective, it is worth noting that DART would benefit from further robust evaluation. For instance, one of the previous evaluations was described as a QED, which, by academic standards, may not fully meet the criteria for a high-quality QED. This shows an opportunity to further strengthen the quality of evidence of DART. Staff at NSPCC and scale-up sites would need to have an understanding that the research on DART shows evidence of promise but that more work is required to rigorously assess its impact. This may be challenging due to the promotion of DART based on the evidence currently gathered about its outcomes.

Interpretation

This study aimed to assess the practical feasibility of conducting a successful impact evaluation of the DART programme. This depends on several key aspects of DART, including its referral routes, eligibility criteria, the potential for randomisation in research design (RCT), suitable comparator group construction, access to administrative data, outcome measurement, sample size and methodological acceptability. Four (out of nine) NSPCC Hubs and four scale-up sites (out of 25) were involved in this feasibility study. The study found challenges in a range of key issues, including referral routes and eligibility criteria, and significant resource constraints that, at this stage, would not make it practical to conduct a full-scale evaluation.

In conclusion, the study found that the current number of families accessing DART is not sufficient to support a well-powered impact evaluation. The required increase in the number of families supported by DART would be substantial, and the ad-hoc and, in some cases, limited referral routes into DART pose a challenge, as there is no clear path to increase the sample size to a level required for a feasible impact evaluation. Developing reliable referral pathways is essential to increase the potential sample size for a robust evaluation, either by increasing the number of current referrals to existing sites or by introducing new sites to deliver DART and increasing partnership work with the wider domestic abuse sector. Funding appears to be a significant barrier to such relationships being strengthened, both for the site running DART and for the organisation unable to enter into a partnership for fear of losing its own funding. Support to build and negotiate pathways could be facilitated centrally by the NSPCCy.

Additionally, the ambiguity surrounding suitability criteria creates challenges in predicting sample size for an impact evaluation. While the use of professional judgment is essential in the domestic abuse sector and valuable in terms of providing person-centred support, a clearer definition of “readiness” or what constitutes the “right” time for families to access DART may be necessary. In addition, simplifying the assessment process for NSPCC Hub staff to mirror scale-up sites could reduce the resource intensiveness of the programme and, therefore, the evaluation. The study also found that an RCT design might be more manageable than a QED approach, but there are still some challenges to this design. First, it would need to be agreed whether the assessment process is considered part of the DART programme and, therefore, whether randomisation should take place before or after assessment. In addition, an RCT approach is not currently favoured by staff members and referrers. Given the widespread, though not universal, concern about RCTs, focusing on building buy-in at all organisational levels and addressing concerns while emphasising long-term benefits is an important step to take. Building buy-in is likely to take some time and would need to emphasise tangible benefits for the families DART works with. Leveraging existing waitlists could mean that staff find an RCT design less objectionable.

In the case of a QED approach, a large comparator group is necessary, and significant resources would need to be allocated to identifying a good source of a comparator group. This would likely involve the need to build relationships with national and local organisations that can provide a comparator group.

More broadly, there were mixed views across the NSPCC and scale-up sites about carrying out an additional evaluation of DART. To encourage support for additional evaluation work, significant extra resources would be needed to facilitate participation among staff. In addition, NSPCC and scale-up site staff would need to understand the additional value an RCT or a QED evaluation could bring beyond what has already been achieved in previous evaluations.

Last, there is a need to improve data quality and management, identify appropriate outcome measures and standardise implementation of them across all sites. A suggestion would be for all sites to adopt the validated scales NSPCC uses to measure the two outcomes (self-esteem measured through the Rosenberg Self-Esteem Scale and attitudes and behaviours measured through the SDQ). It may also be worth conducting cognitive testing of the identified scales to ensure they are appropriate for working with survivors of domestic abuse. Alongside this, it would be important to identify a validated scale for the quality of the mother–child outcome. If a design were planning to incorporate administrative data or integrate data collection as part of a site's standard engagement with participants, it would be necessary to invest significant resources in improving data quality.

Currently, neither an RCT nor a QED design is practically feasible due to the substantial resources required for increasing referrals and delivery, addressing administrative data and outcome measurement concerns, and the lack of acceptability among staff and referrers. Based on the challenges and resource constraints identified, the study recommends these challenges be addressed before considering a pilot study.

Developing a comprehensive plan, securing additional resources, clarifying eligibility criteria, refining referral pathways and building buy-in for a potential RCT are necessary next steps. Once these issues are resolved, a pilot study or full-scale impact evaluation can be reconsidered, ensuring that any further research is based on a more solid foundation.

Limitations

It is crucial to acknowledge the limitations of this feasibility study. These include:

- **Data quality constraints:** the analysis of administrative data was limited due to inconsistencies across sites and concerns about data accuracy. These limitations resulted in information gaps that prevented the inclusion of all records in the analysis. Consequently, the findings of the quantitative analysis should be interpreted with caution.

- **Lack of diversity in cohorts:** There was difficulty in achieving diversity among DART cohorts, with only four scale-up sites (including one co-delivering with an NSPCC Hub) out of 25 and four hubs out of nine NSPCC Hubs included in the qualitative parts of the study. The main reason for this limitation was the capacity constraints of the sites in engaging with the research, potentially introducing biases.

Despite these limitations, the study has several strengths:

- **Research engagement:** the study explored a wide range of views of varying seniority levels across all sites. This allowed for a more holistic assessment of DART.

- **Robust methodological approach:** despite some challenges, the study used a robust methodological approach when assessing the DART programme, combining quantitative and

qualitative data collection methods, and provided an in-depth understanding of the programme's strengths and areas of improvement on a range of issues.

Implications for the Intervention

- Strengthen referral pathways: the feasibility study found that referral pathways to DART are not robust enough to support a large increase in the number of families accessing the programme. As such, strengthening referral pathways by collaborating with local authority teams, police forces and national domestic abuse organisations to develop more effective referral strategies should be a priority.
- Clarify eligibility criteria and assessment process: the current eligibility or suitability criteria and the assessment process for DART are not clearly defined or consistently applied across sites. The NSPCC could gather all practitioners together and try to develop a shared understanding of this; additionally, the assessment process needs to be simplified for NSPCC Hubs to alleviate some pressure on staff members.
- Improve data quality and implement data management systems: data quality is low across several DART sites, which prevents the ability to conduct a rigorous impact evaluation. There should be a focus on investing some resources to improve data quality across all sites.

Implications for the context

- Build relationships with national and local organisations: the feasibility study highlighted the importance of building relationships with national and local organisations that provide services to families that experience domestic abuse. These relationships could support identifying a suitable comparator group for a QED approach – if this approach were taken. Strengthening relationships and partnership work are also likely to increase the number of families referred to DART.
- Address funding and resource constraints: the feasibility study identified funding and resource constraints as major barriers to the implementation of a full-scale impact evaluation. Developing strategies to secure additional funding and resources to support the development and evaluation of the DART programme will be key.

Implications for future evaluation

- Future research efforts should be geared towards addressing the above challenges, including increasing referrals, sample size and staff capacity, improving data quality, achieving diversity in cohorts, and adopting standardised outcome measures. Once these issues have been adequately addressed, a pilot study or full-scale impact evaluation could be considered.
- Develop a comprehensive plan for addressing identified challenges: a plan should be developed that outlines the resources required to address the challenges identified in this feasibility study. This plan should include timelines, milestones and responsible parties for each task.
- Involve stakeholders in the evaluation process: all relevant stakeholders, including DART staff, programme participants (families) and referral partners, should be involved and consulted in the evaluation process to ensure the evaluation is relevant, meaningful and acceptable to all parties involved.

The feasibility study has provided valuable insights into the feasibility of conducting a rigorous impact evaluation of the DART programme. The findings of the study highlight the need to address several challenges, including increasing referrals, improving data quality and standardising outcome measures. Once these issues have been adequately addressed, a pilot study or full-scale impact evaluation could be considered. By addressing these issues, the DART programme can be put on a path towards rigorous evaluation, allowing for the collection of evidence to assess its potential to benefit families facing domestic abuse.

References

Alejo, K. (2014). Long-term physical and mental health effects of domestic violence. *Research Journal of Justice Studies and Forensic Science*, *2*(5): 82–98.

Anderson, K., & van Ee, E. (2018). Mothers and children exposed to intimate partner violence: a review of treatment interventions. *International Journal of Environmental Research and Public Health, 15*(9): 2–25.

Austin, A. E., Shanahan, M. E., Barrios, Y. V., & Macy, R. J. (2019). A systematic review of interventions for women parenting in the context of intimate partner violence. *Trauma, Violence, & Abuse, 20*(4): 498–519.

Bowen, E. (2017). Conduct disorder symptoms in pre-school children exposed to intimate partner violence: gender differences in risk and resilience. *Journal of Child & Adolescent Trauma*, *10*: 97–107.

Callaghan J., Alexander J., & Lisa C. F. (2016). Children's embodied experience of living with domestic violence: "I'd go into my panic, and shake, really bad". *Subjectivity, 9* (4): 399–419.

Carpiano, R. M. (2002). Long roads and tall mountains: the impact of motherhood on the recovery and health of domestic abuse survivors. *Health Care for Women International, 23*(5): 442–449.

Coburn, A., & Gormally, S. (2014). 'They know what you are going through': a service response to young people who have experienced the impact of domestic abuse. *Journal of Youth Studies, 17*(5): 642–663.

Craig, P.; Dieppe, P.; Macintyre, S.; Michie, S.; Nazareth, I.; Petticrew, M. Developing and evaluating complexinterventions: The new Medical Research Council guidance. BMJ 2008,337, a1655.

Dodd, L. (2009). Therapeutic groupwork with young children and mothers who have experienced domestic abuse. *Educational Psychology in Practice*, *25*(1): 21–36.

Domestic Abuse Intervention Programs. (2017). *Understanding the power and control wheel.* Available at: https://www.theduluthmodel.org/wheels/faqs-about-the-wheels/.

Domoney, J., Fulton, E., & Stanley, N. (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*: 539–551.

Donagh, B. (2020). From unnoticed to invisible: the impact of COVID-19 on children and young people experiencing domestic violence and abuse. *Child Abuse Review, 29*: 387–391.

EIF. (2021). *Improving services for children affected by domestic abuse*. Available at: https://www.eif.org.uk/report/improving-services-for-children-affected-by-domestic-abuse.

EIF. (2023). *Learning together about how best to support children and families affected by, or at risk of, domestic abuse.* Available at: https://www.eif.org.uk/blog/learning-together-about-how-best-to-support-children-and-families-affected-by-or-at-risk-of-domestic-abuse.

Family Rights Group. (2023). *Core group meeting*. Available at: https://frg.org.uk/get-help-and-advice/what/child-protection/core-group-meeting/.

Flick, U., von Kardorff, E., & Steinke, I. (2004). *A companion to qualitative research*. London: SAGE.

Grip, K., Almqvist, K., & and Broberg, A. G. (2011). Effects of a group-based intervention on psychological health and perceived parenting capacity among mothers exposed to intimate partner violence (IPV): a preliminary study. *Smith College Studies in Social Work*, *81*(1): 81–100.

Galvani, S. (2006). Safety first? The impact of domestic abuse on women's treatment experience. *Journal of Substance Use*, *11*(6): 395–407.

Haeseler, L. A. (2013). Women's coping experiences in the spectrum of domestic violence abuse. *Journal of Evidence-Based Social Work, 10*(1): 33–43.

Hine, B., Bates, E. A., & Wallace, S. (2022). "I have guys call me and say 'I can't be the victim of domestic abuse'": Exploring the experiences of telephone support providers for male victims of domestic violence and abuse. *Journal of Interpersonal Violence, 37*(7–8): NP5594–NP5625.

Home Office. (2019). *The economic and social costs of domestic abuse*. Available at: https://www.basw.co.uk/system/files/resources/costs_domestic_abuse.pdf.

Howarth E., Moore T. H. M., Welton N. J., Lewis N., Stanley, N., & MacMillan, H. (2016). IMPRoving Outcomes for children exposed to domestic ViolencE (IMPROVE): an evidence synthesis. *Public Health Research, 4*(10).

Humbert, T.K., Engleman, K., & Miller, C. E. (2014). Exploring women's expectations of recovery from intimate partner violence: a phenomenological study. *Occupational Therapy in Mental Health, 30*(4): 358–380.

Humphreys C., Mullender, A., Thaira, R., & Skamballis, A. (2006). Talking to my mum: developing communication between mothers and children in the aftermath of domestic violence. *Journal of Social Work, 6*: 53–63.

Humphreys, C., Houghton, C., & Ellis, J. (2008). *Literature review: Better outcomes for children and young people affected by domestic abuse – directions for good practice*. Scottish Government: Edinburgh. Available at: http://dera.ioe.ac.uk/9525/1/0064117.pdf.

Javaherian, H., Krabacher, V., Andriacco, K., & German, D. (2007). Surviving domestic violence: rebuilding one's life. *Occupational Therapy In Health Care, 21*(3): 35–59.

Kilpatrick K. L., & Williams, L. M. (1998). Potential mediators of post-traumatic stress disorder in child witnesses to domestic violence. *Child Abuse & Neglect, 22*(4): 319–30.

Leech, N. L., & Onwuegbuzie, A. J. (2007). An array of qualitative data analysis tools: a call for data analysis triangulation. *School Psychology Quarterly*, *22*(4): 557–584.

Lee, H., & Munk, T. (2008). Using regression discontinuity design for program evaluation. Proceedings of the Survey Research Methods Section, American Statistical Association

Macintyre, S. (2011). Good intentions and received wisdom are not good enough: The need for controlled trials in public health. *Journal of Epidemiology & Community Health 65*: 564–567. https://jech.bmj.com/content/65/7/564.info

Maikovich, A. K., Jaffee, S. R., Odgers, C. L., & Gallop, R. (2008). Effects of family violence on psychopathology symptoms in children previously exposed to maltreatment. *Child Development, 79*: 1498–1512.

McKenzie, L. (2017b). 'It's not ideal': Reconsidering 'anger' and 'apathy' in the Brexit vote among an invisible working class. *Competition & Change*, *21*(3): 199–210.

Meltzer, H., Doos, L., Vostanis, P., Ford, T., & Goodman, R. (2009). The mental health of children who witness domestic violence. *Child & Family Social Work, 14*: 491–501.

Mullender, A., Hague, G., Imam, U. F., Kelly, L., Malos, E., & Regan, L. (2002). *Children's perspectives on domestic violence*. London: Sage.

NatCen. (2020). *Framework analysis in NVivo*. Available at: https://stagingnatweb.natcen.ac.uk/our-expertise/methods-expertise/qualitative/framework/.

Newing, H., Eagle, C., Puri, R., & Watson, C. W. (2010). *Conducting research in conservation: Social science methods and practice*. London: Routledge.

NIHR. (2019). *Domestic violence and abuse: How NIHR research is helping families in the South West*. Available at: https://arc-swp.nihr.ac.uk/news/nihr-research-domestic-violence-and-abuse/.

NSPCC. (2021). *Domestic Abuse, Recovering Together (DART) service delivery manual*.

NSPCC. (2023). *The evidence base*. Available at: https://learning.nspcc.org.uk/services-children-families/dart.

ONS. (2023). *Domestic abuse victim characteristics, England and Wales: Year ending March 2023.*

ONS. (2022). *Domestic abuse victim characteristics, England and Wales: Year ending March 2022*

ONS. (2020). *Domestic abuse during the coronavirus (COVID-19) pandemic, England and Wales*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/domesticabuseduringthecoronaviruscovid19pandemicenglandandwales/november2020#main-points

ONS. (2012). *2011 census: key statistics for local authorities in England and Wales.* Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuskeystatisticsforlocalauthoritiesinenglandandwales .

ONS. (2022a). *Dataset: Domestic abuse prevalence and victim characteristics*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/domesticabuseprevalenceandvictimcharacteristicsappendixtables.

ONS. (2022b). *Dataset: Domestic abuse in England and Wales – data tool.* Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/domesticabuseinenglandandwalesdatatool

ONS. (2022c). *Census 2021: Ethnic Group.* Available at: https://www.ons.gov.uk/datasets/TS021/editions/2021/versions/1.

Perry, S., & Frampton, I. (2018). Measuring the effectiveness of individual therapy on the well-being of children and young people who have experienced abusive relationships, particularly domestic violence: a case study. *Counselling and Psychotherapy Research, 18*: 356–368.

Refuge. (2021). *Covid service review*. Available at: https://refuge.org.uk/wp-content/uploads/2021/03/Refuge-Covid-Service-Report.pdf.

Ritchie, J., & Lewis, J. (eds.) (2003). *Qualitative research practice*. London: SAGE.

Secco, L., Letourneau, N., & Collins, E. (2016). 'My eyes were open': awakened maternal identity and leaving violent relationships for the infant/children. *Journal of Family Violence*, *31*: 639–645.

Smith, E. (2016). *Domestic Abuse, Recovering Together (DART) evaluation report*. Available at: https://library.nspcc.org.uk/HeritageScripts/Hapi.dll/filetransfer/2016DomesticAbuseRecoveringTogetherDART.pdf?filename=CC18C70DB7C8C3D49403BB94EB176F95207E5F66235DCA89651F5ED2BA5DA9311A3547010EB17451D2DDDA019569BD581EA0CD5852636BDD968745307585651128D12FFD3F139432B70DA68E7496F27BF44D214CFD477CB6F07D22056BB4EAEED59D8D0E983741BD3AAC39728F24B7260DB6C2D1B05918CF263D1C6DB550CF6B034DCC&DataSetName=LIVEDATA.

Smith, E., Belton, E., Barnard, M., Fisher, H. L., & Taylor, T. (2015). Strengthening the mother-child relationship following domestic abuse: service evaluation. *Child Abuse Review*, *24*: 261–273.

Smith, E., Belton, E., & Cooke, S. (2020). *Impact evaluation of the scale-up of Domestic Abuse, Recovering Together.* Available at: https://learning.nspcc.org.uk/media/2356/impact-evaluation-scale-up-domestic-abuse-recovering-together.pdf.

Straus, M. A., & Gelles, R. J. (1990). *Physical violence in American families: Risk factors and adaptations to violence in 8,145 families*. New Brunswick, NJ: Transaction.

Sweet, D., Winter, K., Neeson, L., & Connolly, P. (2020). Assessing the reliability and validity of an outcomes star. Journal of Children's Services, 15(3): 109–122. https://doi.org/10.1108/JCS-03-2020-0009

Trabold, N., McMahon, J., Alsobrooks, S., Whitney, S., & Mittal, M. (2020). A systematic review of intimate partner violence interventions: state of the field and implications for practitioners. *Trauma, Violence, & Abuse*, *21*(2): 311–325.

Walklate, L., Godfrey, B., & Richardson, J. (2022). Changes and continuities in police responses to domestic abuse in England and Wales during the Covid-19 'lockdown'. *Policing and Society*, *32*(2): 221–233.

White, H. (2019). The twenty-first century experimenting society: the four waves of the evidence revolution. *Palgrave Communication, 5*(1): 47. https://doi.org/10.1057/s41599-019-0253-6

Appendix A

INFORMATION SHEET FOR CASE STUDY SITE STAFF

Ethical Clearance Reference Number: LRS/DP-22/23-35596

Title of research

DART Feasibility study

Invitation Paragraph

We would like to invite you to participate in the DART Feasibility study. Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information.

Who will conduct the research?

The research is being carried out by researchers from the Policy Institute at King's University. The research has been commissioned by and is funded by the Youth Endowment Fund (YEF)- https://youthendowmentfund.org.uk/

What is the purpose of the research?

We are interested to learn more about the DART programme and whether it would be possible to carry out an effective impact study of the programme. To do this we are working with sites currently delivering, or preparing to deliver the DART programme to better understand how they deliver DART, and wish to explore with them what the most effective approaches to conduct an impact study of the DART programme.

Why have I been invited to take part?

You are being invited to participate in this research because you work at a DART site.

What will happen if I take part?

If you choose to take part in the research you will be asked to take part in an interview with one of our researchers. We are planning to visit your site as part of the research, and so ideally would carry this out face-to-face, and the interview will last about 45-60 minutes. However, if more convenient we can also carry out the interview online or over the phone.

During the interview we will ask you about your experience of delivering the DART programme, including the referral process, the assessment process, delivering the sessions, and what impact you think the programme has. **Please do not discuss details of specific cases but focus on general themes instead.** We will also discuss with you some ways we could conduct an impact evaluation, and ask for your opinion on them.

Additionally to the interviews, we are planning to carry out observations of some team planning sessions. This would involve a researcher attending one of your sessions and taking notes. This

will help us understand how DART is delivered in practice, and any issues you face that would be relevant to an impact study.

Access for support

The interview may touch upon distressing subjects. If, during the interview, you experience distress, please refer to the below organisations to access support if required:

- **Samaritans -** https://www.samaritans.org/
- **Citizens Advice** - https://www.citizensadvice.org.uk/
- **MIND** - https://www.mind.org.uk/
- **Mental Health Foundation** - https://www.mentalhealth.org.uk/

The following organisations provide support for children, adults and families, specifically:

- **NSPCC Helpline** - If you're worried about a child, or you work with children and need advice or information, you can contact the NSPCC Helpline on 0808 800 5000.
- **NSPCC Childline -** Children and young people can contact the NSPCC's Childline service free and confidentially at any time, by phone on 0800 1111 or via the Childline website www.childline.org.uk
- **The National Association for People Abused in Childhood –** NAPAC offers support to adult survivors of abuse and training for those who support them. NAPAC can be contacted on 0808 801 0331 (calls will not show on your bill) or you can visit their website, https://napac.org.uk/
- **Family Lives –** Offers a free and confidential helpline service for families. You can call them on **0808 800 2222** about any aspect of parenting and family life or visit their website - https://www.familylives.org.uk/

Do I have to take part?

Participation is completely voluntary. You should only take part if you want to and choosing not to take part will not disadvantage you in any way. Your employer will not be made aware of your participation or individual responses.

Once you have read the information sheet, please contact us if you have any questions that will help you make a decision about taking part. If you decide to take part we will ask you for verbal consent at the beginning of the interview.

Data handling and confidentiality

Your data will be processed under the terms of UK data protection law (including the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018)

King's University is the Data Controller for this research. This means we are responsible for making sure your personal information is kept secure and confidential. The research team at King's will anonymise all data. Your name and any other identifiable information will be removed. The interview will be recorded and transcribed. The information you give us will be held securely on a King's College London server, and only staff working directly on the research will be able to access it. Your name and contact details will be kept separately from your interview and observation responses.

We may share the recording of your interview with a transcription service. Data Sharing Agreements will be in place before we share any information about you, and data will be transferred securely via secure file transfer.

The research will end in November 2023. We will retain your personal data until 30th November 2023 unless you withdraw from the research. After this date, we will destroy all personal data we hold as part of the research.

Potential disclosures

If, during the research, you disclose information about safeguarding concerns (such as child or adult abuse), we may have a legal obligation to report this, and confidentiality may be suspended to ensure safety to yourself or others, so that relevant support services can be contacted.

Data Protection Statement

If you would like more information about how your data will be processed under the terms of UK data protection laws please visit the link below:

https://www.kcl.ac.uk/research/support/research-ethics/kings-college-london-statement-on-use-of-personal-data-in-research

What if I change my mind about taking part?

You are free withdraw at any point of the research without having to give a reason. Withdrawing from the reswill not affect you in any way. You are able to withdraw your data from the research up until 30th June 2023, after which withdrawal of your data will no longer be possible as it will have been analysed and combined with other data for the report. If you choose to withdraw from the research we will not retain the information you have given thus far.

How is the research being funded?

This research is being funded by the Youth Endowment Fund. Their website can be found here: https://youthendowmentfund.org.uk/

What will happen to the results of the research?

The results of the research will be summarised in a report for the Youth Endowment Fund. However, we won't use your name and we will ensure any quotes we use wo't identify who said what. This report may be published on the Youth Endowment Fund's website and may be used in other publications.

Who should I contact for further information?

If you have any questions or require more information about this research, please contact me using the following contact details:

Hannah Piggott

Email: edit@kcl.ac.uk

Address: The Policy Institute, King's College London, Virginia Woolf Building, 22 Kingsway, London, WC2B 6LE

What if I have further questions, or if something goes wrong?

If this research has harmed you in any way or if you wish to make a complaint about the conduct of the research you can contact King's College London using the details below for further advice and information:

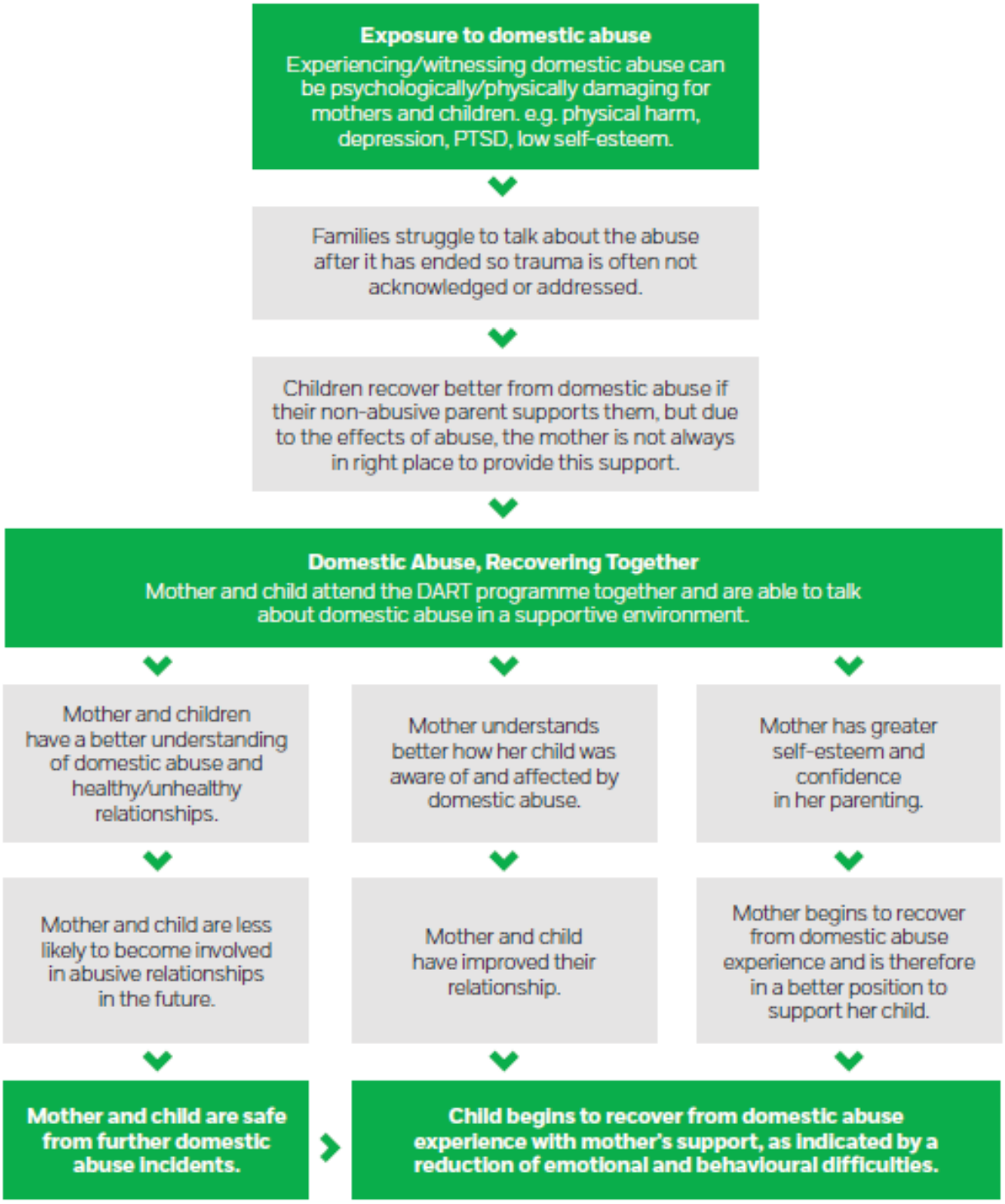
The Chair, Social Science & Public Policy, Humanities and Law RESC, *rec@kcl.ac.uk*

If you would like to complain about the research you can talk to any of the NSPCC staff you know. Or you can phone the NSPCC on 020 7825 2505 (please say that you are calling about the DART Feasibility research research). If you prefer, you can email *researchcomplaints@nspcc.org.uk*

Thank you for reading this information sheet and for considering taking part in this research.

Appendix B

DART Theory of Change



This Theory of Change was developed by the NSPCC DART Team, and comes from their 'DART Training: Domestic Abuse, Recovering Together' document.