

# Supporting Good Judgement in Social Work

<b>Intervention Developer</b>	Cardiff University and UCL School of Management
<b>Delivery Organisations</b>	Cardiff University
<b>Evaluator</b>	What Works Centre for Children's Social Care
<b>Principal Investigator</b>	David Wilkins
<b>Protocol Author(s)</b>	David Wilkins, Catherine Foster, Michael Sanders and Louise Reid
<b>Type of Trial</b>	Lab-In-The-Field RCT
<b>Age or Status of Participants</b>	Social Workers - all aged above 18
<b>Primary Outcome(s)</b>	Brier scores (accuracy of forecasts)
<b>Secondary Outcome(s)</b>	Forecast accuracy distance measure
<b>Contextual Factors</b>	Local authority, professional characteristics of participants

## Table of Contents

Background and Problem Statement .....	3
Intervention and Theory of Change .....	4
Impact Evaluation .....	6
<i>Primary Research Questions</i> .....	7
<i>Design</i> .....	7
<i>Randomisation</i> .....	9
<i>Participants</i> .....	9
<i>Sample size/Minimum Detectable Effect Size (MDES) Calculations</i> .....	10
<i>Main Outcome Measures</i> .....	11
<i>Exploratory Outcome Measures</i> .....	11
<i>Analysis Plan</i> .....	11
Ethics & Participation.....	12
Registration .....	12
Data protection .....	12
Personnel .....	12
Timeline.....	13

## Background and Problem Statement

Child and family social workers in England make thousands of influential decisions every day and by so doing they aim to support families and protect children from significant harm. Deciding that a child is at risk of significant harm is one of the most consequential decisions that the state can make, leading as it may to the child being removed from the care of their family, permanently in some cases. Less critically, but still of importance, social workers must also make decisions about whether parents might benefit from particular services, whether to close or keep open a case, how often to visit children and what information to include in assessment reports. It is widely acknowledged that the ecology of social work decision-making is complicated. Social workers do not make all of their decisions individually. In many cases, good decision-making means working in partnership with parents and young people and enabling families to take decisions for themselves. Decision-making in social work is also informed by inequalities at the macro level, by institutional and organisational factors, including the availability (or otherwise) of resources, specific local and team cultures and family-related factors (Keddell and Hyslop, 2016). In addition, social work is not simply a professional and evidence-informed activity – it is to a great extent a moral and political activity too. And just because a particular outcome is thought likely, does not imply a necessary course of action – *is* does not imply *ought*. The challenge often in social work is to balance competing rights. For example, parents are entitled to support and the opportunity to show they can make changes, even if the chances of success are deemed (by the social worker) to be low. On the other hand, children are entitled to protection from significant harm and the right of the parent to support does not supersede the child's right to protection. Thus, understanding and supporting the quality of social work decision is far from straight-forward. In order to ensure the system as a whole consistently does the right thing with and for families, changes would be needed at all levels. This complexity does not mean however that the role and influence of the individual social worker is unimportant. Just as it is worthwhile and valuable to support social workers to develop excellent communication skills (despite all of the wider macro influences alluded to above), so it is worthwhile and valuable to support social workers to make the best possible decisions – wise, professionally-informed, child and family focused, consistent, reasonable and fair.

There are two broad approaches that are used to assess the quality of decision-making - by measuring the outcome, or by measuring some aspect of the process. Neither approach is straight-forward, and both have strengths and limitations. For example, when measuring the outcome, one has to decide over what time-frame to consider. If a social worker decides to close a case they are working on, is it reasonable to judge the outcome of this decision over weeks, months or years? If the case is closed and the child is significantly harmed within a short-period of time, one might judge the decision to be deficient in some way. One difficulty with measuring outcomes is that good decision-making may lead to poor outcomes because of other factors (and vice versa) – including the influence of chance. It may also be the case that more expert decision-makers are given more complicated decisions (e.g. senior social workers being allocated more complex child protection work). This could give the impression that as expertise increases, the likelihood of a more positive outcome following a decision decreases.

Measuring the quality of the process is also complicated, and usually involves either measures of coherence or correspondence. Coherence measures include asking whether the rationale for the decision is reasonable and defensible. For social work, one might also consider to what extent the decision was made in collaboration with family members, as opposed to by professionals without any family involvement. Correspondence measures consider whether the process conforms to some external standard, for example of acknowledged experts. Novice social workers can be asked to consider a case vignette and make a decision about what should happen next. The rationale for these decisions can then be compared with those of experts considering the same vignette (see Taylor and Whittaker, 2018).

Another approach is that used by the *Good Judgement Project*. In a series of studies, they have assessed the quality of expert reasoning and judgement by considering correspondence not between experts and but between forecasts and reality. They do so by asking respondents questions about what might happen in future and using Brier scores to keep track of accuracy. By so doing, they have been able to identify some common features of good forecasters, some training approaches to help improve forecasting accuracy and ways of supporting teams to work together to produce more accurate forecasts than those of individuals. Many of the approaches they have used will be familiar to social work – including the importance of critical thinking, keeping an open mind, avoiding cognitive biases and considering a diversity of opinion.

The relationship between forecasting and decision-making is not straight-forward, although it is true that planning in general relies on an ability to think about what might happen (or not happen) in future. In social work, for example, when completing assessments social workers are making a forecast about the future (however implicitly or explicitly) – what could happen if we do x and what could happen if we do y. If a child or family presents with these needs, in the context of z, what are the possible and probable outcomes if the local authority a) intervenes or b) does not intervene? Yet there are all sorts of real-world pressures that impact on the kinds of plans (and underlying forecasts) that people make. One example in social work is the consequence of being wrong about a child being significantly harmed. In practice, it is often preferable (for the system, if not for families) that workers err on the side of caution (to be risk-averse). Thus, some plans and forecasts may be made not with the intention of achieving accuracy, but because of other motivations.

With this study, we want to explore the concept of forecasting in relation to social work practice. The study has three aims - i) explore whether social workers can answer forecasting questions in relation to short case vignettes, ii) measure whether these forecasts are on average more accurate than chance and iii) to see whether brief online interventions make a difference (whether they improve accuracy). The outcome of this study, if successful, would be further studies involving social workers, to further explore these ideas.

## Intervention and Theory of Change

Based on the experience of the *Good Judgement Project*, we hypothesise that forecasting accuracy can be improved by using relatively brief on-line interventions. For this experiment, we have selected three such interventions. Participants will be randomly assigned to one of these three interventions or a control condition. The three interventions are as follows.

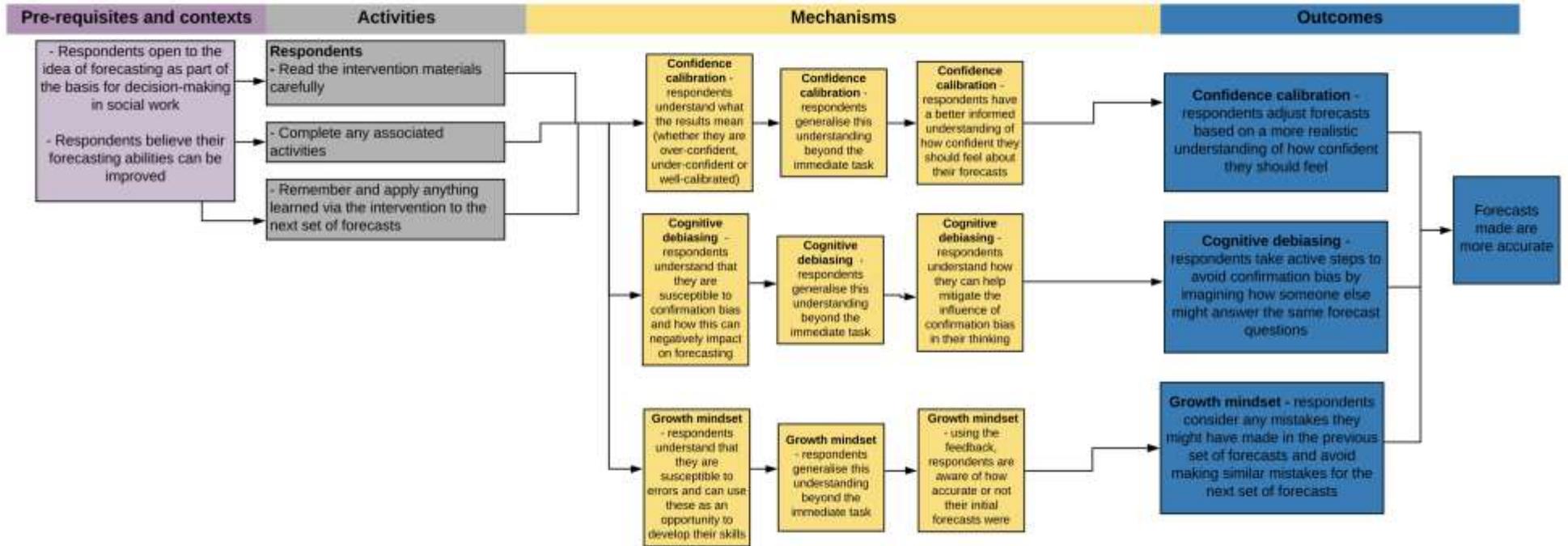
**Cognitive de-biasing:** From the work of Kahneman and others, it has long been recognised that human reasoning is subject to a number of systematic errors. For example, confirmation bias occurs when people pay more attention and give more weight to evidence that supports their current view and overlook or undermine evidence which supports a different point of view. Confirmation bias has also been referred to as ‘my side’ bias. Cognitive de-biasing is a brief intervention that aims to make people more aware of confirmation bias and provides a framework for actively seeking out reasons why an initial conclusion may not be correct. This intervention will involve participants reading some material about confirmation bias and being given some advice about steps they can take to help minimise it (Morewedge et al., 2015; Chang et al, 2016).

**Feedback and growth mindset:** Providing people with feedback on their mistakes (and successes) is a well-established method for improving performance in a range of different situations and contexts (Wrisberg, 2007). For this intervention, participants will be given feedback on how they performed in relation to their first set of forecasts.

**Confidence calibration:** Understanding how confident one should feel about one’s level of expertise is a challenge for many people. Particularly for novices and near-experts, it can be hard to judge your

own expertise in relation to others. For this intervention, participants will be given a series of social work-related general knowledge questions and asked to give a true / false answer and an indication of how confident they feel. At the end, participants will be shown whether they are over or under-confident or whether they are well calibrated (with confidence levels roughly equivalent to their level of knowledge; Hattie, 2013).

The rationale for all three arms is that relatively short interventions can impact on reasoning abilities, at least in the short-term (Tetlock and Gardner, 2016).



## Impact Evaluation

### Primary Research Questions

- How accurately can social workers forecast what is going to happen next from reading brief referral-type case vignettes?
- Can the accuracy of social work forecasting be improved using short, online interventions?

### Exploratory Research Questions

- Can we predict forecasting accuracy from personal or professional characteristics?
- Is social workers' forecasting accuracy more similar to others in the same local authority than to the wider sample and / or actual case outcomes (this question seeks to explore the question of different thresholds between authorities and to what extent these influence forecasts about what might happen next)?

### Design

<b>Trial type and number of arms</b>		Randomised controlled trial with four arms (three intervention and one control group)
<b>Unit of randomisation</b>		Individual social workers
<b>Primary outcome</b>	variable	Accuracy of forecasts
	measure (instrument, scale)	Brier scores
<b>Secondary outcome(s)</b>	variable(s)	Accuracy of forecasts
	measure(s) (instrument, scale)	Distance measure
	variable(s)	Accuracy of forecasts
	measure(s) (instrument, scale)	Brier score differences between timepoints (pre and post training) and each training intervention

For this experiment, we will ask social workers to forecast the likely outcome of a series of cases, having been provided with 300 – 500-word summaries of the case from the point of initial referral. Each case example is drawn from a real-world situation, and so we already know what actually happened in relation to each of the questions posed. For each case example, respondents will be asked to judge the likelihood of each of the following outcomes:

- No further action
- Social work or other form of assessment
- Emergency removal into care
- Something Else

Respondents will also be asked to judge the likelihood of the child becoming subject to the following types of plan within 12 weeks of the referral:

- Child in Need plan (section 17, 1989 Children Act)
- Child protection plan (section 47, 1989 Children Act)
- Looked after child plan (section 20 / 31, 1989 Children Act)
- No plan

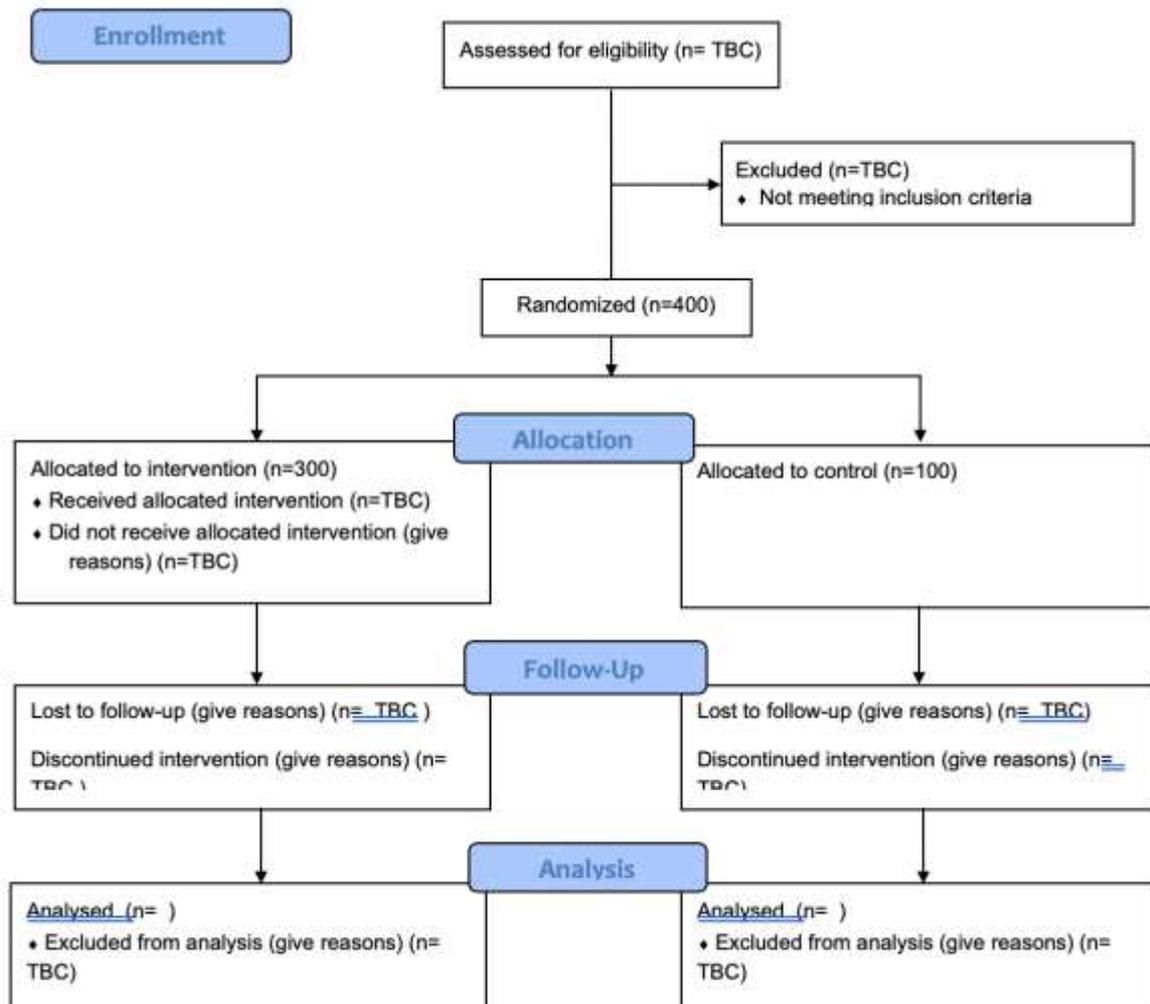
The rationale for using a 12-week period is that when considering a referral, our hypothesis is that decisions are made at that point with a view to a relatively short time-window. It may be that forecasts over the longer term are actually easier, depending on what is being forecast.

We will also ask additional questions about each referral in relation to the possible behaviour of family members (e.g. will the parent attend a parenting programme or will the parent meet with the social worker?).

We have chosen to use case studies at the point of referral because information is often relatively limited at this point in time, and so respondents will not have to read a great deal of information in order to take part (which is then somewhat reflective of reality as this point within the referral and assessment process, albeit no laboratory test can be considered directly comparable to the real-world experience of responding to referrals).

After completing two initial forecasting tasks, respondents will be randomly allocated into one of four conditions - control or three forms of intervention.

- In each intervention arm, respondents will take part in a short (<10 minutes) activity with the aim of improving their forecasting ability - cognitive debiasing, confidence calibration and feedback with a growth mindset.
- After the intervention activity, respondents will complete a further set of forecasting tasks.
- Finally, each respondent will be asked to complete an endline survey to gather data on personal and professional characteristics.
- We will then compare the accuracy of the forecasts made by respondents in each of the four conditions before and after the intervention.



## Randomisation

- Participants will be individually allocated to either a control group or one of three intervention groups.
- Randomisation will be undertaken blindly using the Qualtrics online survey website allocating equal numbers to each condition.
- The randomisation process will be recorded by Qualtrics and will form part of the data output from the completed surveys.
- Analysts will be blinded to group allocation.

## Participants

- All registered social workers working for local authorities in England are eligible to take part.
- The study will be advertised using the What Works for Social Care website and social media platforms.
- We will also approach partner local authorities and ask them to publicise the study within their departments. We may also approach other social work organisations in England (e.g. Frontline) and ask them to do the same.

## Sample size/Minimum Detectable Effect Size (MDES) Calculations

		MDES (Proportion of a Standard Deviation)
<b>Alpha</b>		0.05
<b>Power</b>		0.8
<b>One-sided or two-sided?</b>		Two-sided
<b>Sample Size (Social Workers)</b>	Intervention	300
	Control	100
	<b>Total</b>	400

For our primary outcome, the aim is to establish a baseline measure of forecasting accuracy, and distance from accuracy measure respectively, for social workers in England. For a 95% confidence level and 5% margin of error a sample size of 383 is considered representative of the population.

An updated power calculation was conducted when “Sample Size Tables for Clinical Studies (Third Ed.) software, capable of estimating sample size for trials with more than two arms, became available to the team. The programme was used to estimate the sample size need to achieve power of 0.8 based on a control vs intervention group 1 post-training mean difference of 0.08 and a common standard deviation of 0.17. The standard deviation was generated from the baseline data of the first 200 respondents and the mean difference of 0.08 was deemed a meaningful difference between groups (see output below).

We will aim to recruit 600 participants using a 1:3 control to intervention ratio. We anticipate that a sizeable number of participants will not complete the survey adequately and estimate that 600 participants will be sufficient to achieve the required sample size of 400 needed to achieve power of 0.8 based on 99 respondents per group.

Sample Size Tables for Clinical Studies - [Several Comparisons with Placebo - Continuous]

Sample Size Calculator    Tabulations    Design Implementation    Tools    Window    Help

**Several Comparisons with Placebo - Continuous**

Significance Level,  $\alpha$     0.05    to       by      

Power,  $1-\beta$     0.6    to    0.9    by    0.1

Difference in Means,  $D_m$     0.08    to       by      

Standard Deviation,  $\sigma$     0.17    to       by      

Number of treatments and placebo,  $g$     4

Two-sided  $\alpha$    

Compute

Several Comparisons with Placebo - Continuous

Each cell gives the number of subjects for the placebo group,  $m$ .  
The number of subjects for each treatment groups is  $n$ , and the total sample size is  $N$ , where  $N = m + n(g-1) = m(1 + \sqrt{g-1})$

$D_m$	$\sigma$	$\alpha$	Power, $1-\beta$			
		Two-sided	0.60	0.70	0.80	0.90
0.08	0.17	0.05	62	78	99	131

Print

## Main Outcome Measures

The main outcome measures are;

- Overall forecasting accuracy measured via pooled pre and post intervention Brier scores.
- Forecasting accuracy changes (Brier score changes) for each intervention type

## Exploratory Outcome Measures

- Similarity between forecasts made by respondents in the same local authority vs. respondents in other local authorities.
- Predictive power of professional and personal characteristics for forecasting accuracy
- Whether certain questions are harder to accurately forecast than others (e.g. comparing process-related questions with behaviour-related questions).

## Analysis Plan

### Primary Analysis

The first step of the analysis is the calculation of Brier scores for each individual using the following formula:

$$Z = (1 - x)^2 + (0 - y)^2$$

Where  $x$  = the forecast for the outcome that occurs and  $y$  = the forecast for the outcome that does not occur.

Median accuracy across the full sample will be calculated by pooling pre and post intervention data.

To test the effectiveness of each intervention on forecasting accuracy, a multiple regression model will be created with post-test Brier scores from each of the four conditions. Baseline Brier scores will be included as a covariate to control for differences in baseline scores between the conditions.

### Exploratory Analysis

We will explore whether forecasting accuracy is predicted by personal and professional characteristics such as education level and current role.

We are interested in whether forecasting accuracy might be associated with certain professional or personal characteristics. For example, are more experienced social workers able to make more accurate forecasts? As part of the endline survey, we will ask social workers about the following:

- Gender
- Age
- Ethnicity
- Highest educational qualification
- Type of social work qualification (e.g. BA, MA, fast-track)
- Working location (employing local authority)
- Current role
- Team
- Experience in current role
- Length of time since qualification as a SW
- Understanding of key purpose of children's service (proxy for professional value-base)
- Confidence in forecasting ability

We expect to find that the following four characteristics will be predictive of forecasting ability than the others:

- Highest educational qualification and type of SW qualification
- Confidence in forecasting ability
- Current role
- Team

A stepwise linear regression model will examine relationships between forecasting accuracy and the four personal and professional characteristics listed above. A final stepwise regression will examine the remaining eight factors that we do not expect to be meaningfully related to forecasting accuracy, but will test in the interests of completion.

## Contextual Factors Analysis

The most significant contextual factor is likely to be the local authority within which each respondent is working. We know that different local authorities operate different thresholds, therefore the social worker may answer correctly for what would happen in their local authority, but incorrectly in relation to what actually happened (in a different local authority). As part of the endline survey, we will ask participants to tell us the first half of their workplace postcode, which should enable us to map participants by local authority. We will examine whether respondents are more consistent with other respondents from the same local authority than they are with the wider sample and actual case outcomes. Whether social workers are able to make accurate forecasts based on brief referral-type information (as in this study) may bear little direct relation to real-world decision-making (or forecasting ability in relation to actual families).

## Ethics & Participation

- Ethical approval has been obtained from the ethics committee of the School of Social Sciences, Cardiff University.
- The trial will be advertised using the What Works for Children's Social Care website and social media platforms. Social workers will be invited to take part by completing a survey, hosted on Qualtrics. We will also ask current partner local authorities to help promote the survey locally and we may also ask other social work organisations (e.g. Firstline) to do the same.
- The first part of the survey will consist of an information sheet and consent form. To proceed with the survey, social workers will need to consent to taking part.
- If any participants decline consent, they will be exited from the survey at that point.
- Participants will be able to exit the study at any stage by exiting the survey, although any data provided at that stage will be retained.

## Registration

- The trial was first registered on the Open Science Framework (OSF) on August 14th 2019. The project details are under embargo until October 31<sup>st</sup> 2019, however the details can be previewed here; <https://osf.io/rptdz/>.

## Data protection

- We will only collect and process data in order to address our research questions. In all circumstances, the identities of individuals taking part in the study and the data they provide will be kept confidential and will only be used for research purposes. Participants will be informed of their right not to take part in the study, either by not consenting to take part at all or by exiting the survey at any point they so choose.
- Data will be processed only when the data subject has given consent to the processing of his or her personal data for the specific purpose of conducting this trial.
- All data collected will be stored securely on Cardiff University computers.

## Personnel

- David Wilkins, Principal Investigator, Cardiff University
- Catherine Foster, Co-investigator, Research Associate, Cardiff University
- Michael Sanders, Co-investigator, What Works for Children’s Social Care and King’s College London
- Louise Reid, Co-investigator, What Works for Children’s Social Care

## Timeline

- Timetable including specification of who is responsible for completing each task
- Include specific dates or date intervals.

Dates	Activity	Staff responsible/ leading
July 2019	Confirmation of experimental design	David and Michael
10 <sup>th</sup> – 12 <sup>th</sup> July 2019	Development of case studies	David
12 <sup>th</sup> July 2019	Completion of experimental system	Catherine
July 2019	Pilot test of case studies	David and Catherine
From August 1 <sup>st</sup> onwards	Recruitment of social workers	David and Louise
November 2019 / January 2020	Analysis	David and Catherine
February 2020	Reporting	David and Catherine