DEBT ADVICE:

**A SCOPING STUDY
FOR MEASURING OUTCOMES**

A REPORT FOR THE MONEY ADVICE SERVICE

Sara Davies, Yvette Hartfree and Elaine Kempson
APRIL 2016

University of
BRISTOL

Acknowledgements

Authors

Sara Davies is Research Fellow based at the University of Bristol's Personal Finance Research Centre.

Yvette Hartfree is Research Fellow based at the University of Bristol's Personal Finance Research Centre.

Elaine Kempson is an Emeritus Professor at the Personal Finance Research Centre.

Contents

1 Background and methodology

1.1 Background

The Money Advice Service has been the statutory body responsible for coordinating the provision of debt advice throughout the UK since April 2012. In 2015 the Service commissioned the University of Bristol's Personal Finance Research Centre to conduct a scoping study to better understand the feasibility of conducting a longitudinal study of clients of debt advice. This is against a backdrop in which more than eight million people in the UK[1] are estimated to be over-indebted and an ambition to double the proportion of these who seek debt advice. In order to achieve this, the Money Advice Service have identified the need to continue to increase provision of face-to-face debt advice and facilitate a shift towards other channels, where this is appropriate to the needs of clients. Critical to this expansion is a need to understand the extent to which the debt advice sector, via its different channels, is effective and efficient at preventing crisis debt, enabling people to overcome it when it arises, and preventing it from happening again. Specifically, the Money Advice Service are interested in conducting a longitudinal survey that would identify the long-term impacts of debt advice, understand any variance in impacts by demographic, delivery channel and organisation and identify the most effective outcomes that help to ameliorate over-indebtedness. This scoping study was commissioned to inform the optimal design of the subsequent, prospective evaluation by making clear and detailed recommendations for its methodology.

1.2 Recommendations needed to inform the design of a longitudinal study

Some of the key recommendations needed from this scoping study included:

- The duration of the study and what constitutes an appropriate longitudinal perspective in order to measure the most operative outcomes.
- The sample source and strategy for ensuring representativeness of those in debt.
- Subsequent sample recruitment, retention and achieved sample size (overall and of any essential or recommended sub-samples), taking into account the stress and vulnerability many debt advice clients will be experiencing and home movers during the course of the longitudinal research period.
- Appropriate comparison or control groups to provide a counterfactual to outcomes in the absence of debt advice.
- Appropriate data collection methodologies, taking into account, for example, lower internet use among debt advice clients than the population as a whole and potential avoidance of phone calls, mail and other communication methods.
- The ability to distinguish outcomes associated with aspects of service delivery (e.g. channel, organisation) from other (confounding) factors, such as client profiles (including demographics and advice needs).

---

[1] "A Picture of Over-Indebtedness", Money Advice Service, 2016.

1.3   Methodology and aims of scoping study

The overall approach for the scoping study was to conduct a targeted, expert-led literature search combined with a systematic approach to analysing and synthesising the evidence collected. We conducted a review of UK and international evidence on studies which have attempted to measure the impact of debt advice outcomes over time, and the lessons from these; in particular:

- What can be learned methodologically from existing studies?
- Which outcomes have effectively been measured?
- When, following the receipt of debt advice, do outcomes occur – which outcomes can be expected to be achieved in the short-term and which are longer-term outcomes?
- How do debt advice outcomes vary by client profile or by delivery channel?

The second strand of the review focussed on identifying best practice in longitudinal impact evaluation in social policy, exploring questions such as:

- Is a counterfactual (control) group possible and what is the best evaluation design for this?
- How to recruit a sample including opt-in / out processes?
- What is the best survey method to use (e.g. face to face, telephone, online) to maximise the response rate and minimise non-response bias?
- How long can longitudinal samples be maintained for and what are the best approaches for minimising sample attrition?

With all this evidence collated, the review explores different options and recommendations for the most optimal research methodology and timeframe for a longitudinal evaluation of debt advice outcomes.

2   Impact evaluations

In order to understand the longitudinal outcomes of debt advice, the Money Advice Service is seeking to construct an impact evaluation in order to answer the overarching question of all evaluations of this type: 'what is the causal effect of an intervention on an outcome of interest?' (Yoong et al, 2013). The Money Advice Service wish to establish which outcomes, including changes in financial behaviour, financial capability, levels of indebtedness, and changes to mental health, can be attributed to the receipt of debt advice.

It is important to note that while there are research designs that are relatively easy to implement and commonly used, many of these can't identify what changes in outcomes are directly attributable to the intervention alone, and not to other factors in the environment (known as 'confounding' factors). These include:

- Comparisons of outcomes before and after the receipt of debt advice for those who received it. In the absence of a control or comparison group there is no way to tell whether the changes in outcomes resulted from the receipt of debt advice or other factors unconnected to it.
- Comparisons of outcomes between people who seek debt advice and those who don't. In this case the key problem is the self-selecting nature of the sample. The act of seeking advice to help resolve a debt issue is likely to indicate differences in attitude, capability or circumstances; confounding factors that can be very difficult to measure. A design therefore needs to account for this, and to attempt to understand the role that the actual debt advice itself had, as separate to the inclination to seek it out.

Establishing what is the *direct* outcome of an intervention, and what may be the result of other factors is never easy to achieve (Peise, Judkins and Kalton, 2009). To do so one needs to know what would have happened in the absence of the intervention (known as the counterfactual). Ideally one would track the outcomes of the same individuals in two parallel universes: one where they received debt advice and one where they did not. In reality, the best we can do is to compare the outcomes of a group of people who experience the intervention (known as the 'treatment' group) with a second group of people who are identical in all respects but did not receive the intervention (known as the 'control' or 'comparison' group, depending on how they were selected).

A further consideration in designing the impact evaluation of debt advice is the possibility that the behaviour of the people being tracked may be influenced by the fact that they are participating in the research (known as the Hawthorn effect). This would be less of a problem if both treatment and control group were likely to be affected equally. But it is likely that, in this case, the treatment group might be more influenced in this; however, the Hawthorn effect becomes lesser or greater of an influencer depending on design used (see flowchart below).
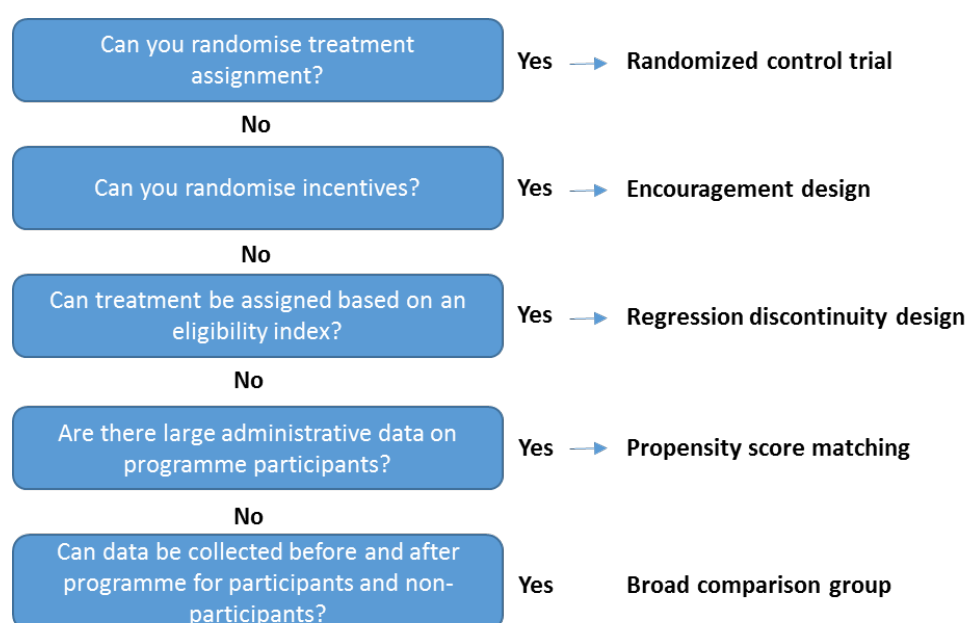
The design of the impact evaluation is further complicated by the fact that there is not one intervention to be assessed, but rather a range of different interventions provided by a number of different agencies, through a number of different channels (face to face, telephone, online) and for different lengths of time and frequency. The types of people seeking advice through different channels might vary in both observed and unobserved ways.

2.1   Identifying a control group

To control for confounding factors the outcomes experienced by the treatment group (in this case recipients of debt advice) need to be compared with the outcomes of a control or comparison group that should:

- Be identical to the treatment group in terms of both observable characteristics (such as level and composition of debt, income, gender, education, etc.) and unobservable characteristics (such as motivation to deal with debt problems, abilities to negotiate with creditors unaided, family support, etc.). This is to ensure that differences in outcomes are not, in fact, the result of differences between the two groups.
- Be expected to react to debt advice in the same way as the treatment group. That means that if the intervention were switched and debt advice provided to the comparison group instead of the treatment group, the impact would be identical.
- Be equally exposed to other interventions as the treatment group. The only difference between the treatment and control groups should be the debt advice intervention being studied.

A flowchart, adapted from Yoong et al (2013 p. 71), summarises the main options possible for the design of an impact evaluation, and some of the factors to consider in the use of these methods:

Can you randomise treatment assignment?	Yes →	Randomized control trial
No		
Can you randomise incentives?	Yes →	Encouragement design
No		
Can treatment be assigned based on an eligibility index?	Yes →	Regression discontinuity design
No		
Are there large administrative data on programme participants?	Yes →	Propensity score matching
No		
Can data be collected before and after programme for participants and non-participants?	Yes	Broad comparison group

Generally, the most robust method of establishing the counterfactual is through experimental methods. The 'gold standard' among these are **randomised control trials**, where a group of people with debt problems is assigned randomly either to receive an intervention, or not. The element of randomisation allows greater inferences to be drawn about the effect that the intervention had over other factors (Peise, Judkins and Kalton, 2009). In reality, ethical concerns may disallow the possibility of using this methodology, because it would involve denying people known to be in financial difficulty access to debt advice, in order to form the control group for the experiment.

Where such concerns exist there are a range of other quasi-experimental methods that can considered. These are presented in the order in which they would normally be considered (as shown in the flow chart above).

- *Randomised encouragement design*. In a study with this design a group of people is identified who are facing debt problems, half of whom are selected at random and encouraged to contact a debt advice service, while the other half is not encouraged to do so. The outcomes of both groups are then tracked. The potential to use this method is discussed later in the report.

Other quasi experimental designs use statistical methods to simulate the effect of randomisation. They include:

- *Regression discontinuity design*. This assumes that there is an eligibility cut-off point for an intervention (for example age) and creating a control group that is just the other side of the cut-off point. An example is the impact evaluation of the Child Trust Fund where children up to the age of five would have been eligible for the Fund, and children aged six were used as the comparison group (and compared with five year olds) (Kempson et al, 2011). The basic idea underpinning this design is that the groups on either side of the cut-off point are very similar. However, it is hard to see how this approach could be used to measure the impact of debt advice.

- *Propensity score matching*. This approach involves creating a comparison group that shares the key characteristics that determine the use of a debt advice service. Although this sounds simple, in practice it is often much harder to achieve a robust comparison group in this way. It is particularly complex where the key characteristics are difficult to measure or may be uncommon among those who have not sought advice (e.g. motivation to sort out debt problems). And if the list of relevant determinants is very large, or if each characteristic takes on many values, it may be hard to identify a match for each person in the treatment group.

- *Broad matched comparison group*. Where none of the above is feasible, it is possible to construct the best possible comparison group, matching the characteristics of the comparison group to those of the treatment group as far as is practicable. This can most easily be done where a service is being rolled out one geographical area at a time, so that the comparison group can be selected from an

area where the service is not yet operational. This is the approach that was used in the evaluation of the Pathfinder Money Guidance Service (Kempson et al, 2010). This is the least satisfactory of the methods considered and provides a broad comparison group only.

Where a randomly assigned control group is not feasible it is possible to improve the assessment of the impact of an intervention by using a difference-in-difference approach to estimate the effect. This involves comparing the extent of change in outcomes between two points in time in the treatment group with the extent of change over the same time period in the control or comparison one. An example might be the change in amounts owed to creditors by both the treatment and the control or comparison group between the period before the treatment group received debt advice and, say, 12 months after receipt of advice. As such it depends upon being able to conduct a survey interview with the treatment and control groups before the treatment group seeks debt advice. This approach was used in both the Child Trust Fund (Kempson et al, 2011) and Money Guidance Pathfinder Evaluations (Kempson et al, 2010) mentioned above.

The less robust the method of selecting the control or comparison group, the greater the need to use a difference-in difference approach to estimating impacts. So it is not needed for a full randomised control trial, but would be highly desirable (if not essential) if a broad comparison group were used.

The final chapter of this report discusses the above options, alongside the other design considerations discussed, in light of the evidence presented in the next two chapters, to arrive at recommendations for the proposed study.

3 Summary of existing debt advice outcomes evaluations

3.1 Methodology of existing evidence

Our review examined reports from studies that included some exploration of the outcomes of debt advice on those who received it. They included some where this was the main focus of the study and some where it was part of a study of wider concerns surrounding debts. They included both qualitative and quantitative studies and ones using mixed methods, and both cross sectional and longitudinal ones. Most were UK-based studies, but five examined US populations. The methodologies and time frames are summarised below.

Method	Timeframe	No.
Qualitative	Cross sectional	7
Qualitative	Longitudinal	3
Quantitative	Cross sectional	5
Quantitative	Longitudinal	7
Mixed method	Cross sectional	5
Mixed method	Longitudinal	1
Literature Reviews		6
Total		34

See appendix for a breakdown of reports or articles included in each category.

3.1.1 Comparison group

We were able to identify only seven studies of debt outcomes that included a comparison group of some kind and, of these, only one employed any form of randomisation, a randomised encouragement design (Pleasance and Balmer, 2007). In this 51 per cent of the 402 participants were offered the opportunity to receive advice. However, the effectiveness of this approach (as is often the case) was diminished as a high proportion of the intervention group did not take up the offer to receive debt advice (73 per cent), while ten per cent of the control group had sought debt advice independently. While use of a randomised encouragement design is a possibility for a debt advice outcomes evaluation, the lack of interest in taking up an offer of debt advice in this study, and in Collard et al (2012) would dictate the need for a sample large enough to withstand a high level of attrition if effects are to be picked up. Collard (2012) failed to recruit the 300 participants that the study originally aimed to achieve. For greater detail on the use of randomised encouragement design and comparison groups see the Pleasance and Balmer case study in section 3.1.7.

The other studies identified had all used a broad comparison group. These studies used a range of methods to select comparison groups of people who had not sought advice:

Study	Sample	Test group	Comparison group
Collard et at (2012)	Housing Association tenants	Housing Association tenants in one particular area who participated in the financial skills training.	Housing association tenants from same housing provider in a different area who were not offered the financial skills training.
YouGov for MAS (2012)	Taken from a YouGov sample of people who had reported being in debt since 2006	Those who reported being in debt and who had sought debt advice.	Those who reported being in debt but had not sought debt advice.
Turley and White (2007)	Sample drawn from the English and Welsh civil justice survey stating had a 'money problem'	Those who said they had sought advice.	Those who said they had not sought advice.
Day et al (2008)	Clients drawn from five partners support agencies	Clients who had used the money outreach service that was being evaluated.	Clients who had used the support agencies, but not the money outreach service.
Elliehausen et al (2003)	Sample provided from credit report data	Details provided by credit counselling services of their clients.	The credit report data of similar borrowers who were not on the list provided by credit counsellors.
Pleasance and Balmer (2007)	Recruited from job centre, people who were in arrears for bills or experiencing difficulty managing debt.	Were offered telephone debt advice via National Debtline (may not have accepted).	Not offered telephone debt advice from National Debtline (may have sought advice independently).
Smith and Patel (2008)	People who used money outreach services	Those who had received advice from the money outreach pilot.	Those who had used other Legal services commission services, but not the money outreach.

There can be little doubt that constructing a robust comparison group will not be easy and we have been unable to find a good example of where this had been achieved. We return to this in the following chapter where we consider a broader range of longitudinal studies.

3.1.2   Sample

In the majority of the studies reviewed, the sample had been recruited via debt advice providers and therefore covered people who had received debt advice. This was typically the only criterion for inclusion in most samples; the level and/or type of debt was generally not used as a sampling criteria (YouGov, 2012 distinguishes between those with manageable and unmanageable debt when sampling), although it was sometimes reported in the analysis. Sample sizes in the quantitative studies ranged from large, such as YouGov, 2012 (4,020) and Optimisa, 2013 (1,902), to as low as 136 (Illuminas, 2008) or 176 in the first wave dropping to 61 by wave three 12 months later (Williams and Samson, 2007). The majority of these did not have a comparison group, however. The studies that used management data for analysis tended to have the largest sample sizes, with thousands of

cases analysed (Smith and Patel, 2008; Elliehausen et al, 2003, 2007). Management data is the data that is collected by an organisation on their customers or clients, which can be collected for a number of reasons, rather than data collected for the purpose of research.

Elliehausen (2003) used the records held by a credit reference agency to identify a broad comparison group. Two studies identified users of other services at the agencies offering debt advice (Day et al, 2008; Smith and Patel, 2008). Another surveyed people living in a contiguous and similar locality (Collard et al 2012).

A small number of studies adopted a different approach altogether and created their samples from people who had reported debt issues or money problems in a previous survey (YouGov, 2012; Turley and White, 2007). In a further case (Pleasance and Balmer, 2007) the sample of people struggling with debt was recruited from users of a job centre.

3.1.3   Defining over-indebtedness:
Regardless of the sampling approach taken, having a clear definition of over-indebtedness would be important for a study in this area involving a control comparison group, since that would be the main criterion for inclusion in the study of both the treatment and the comparison group.

Overall, no single definition of **over-indebtedness** emerged from the studies reviewed, and frequently this concept was not defined in the reports. As already mentioned, often the criterion for inclusion in the studies was self-referral for debt advice. Fitch et al (2009) noted a lack of consistency in the conceptualisation and measurement of debt in the literature while conducting a systematic review of the links between debt and mental health.

Other studies often used a self-assessed measure of over-indebtedness; for example, whether individuals had 'unmanageable debt' (YouGov, 2012) or were facing money problems (Turley and White, 2007). In the second of these not all of those who reported having a money problem actually had problem debts or were in financial difficulties, suggesting that the wording of the screening question was too broad (Turley and White, 2007). London Economics (2012) used a proxy measure of the propensity to fall into financial difficulty: individuals who were spending more than 25 per cent or 50 per cent of their gross monthly income on unsecured repayments. However, this seems too wide for the proposed study.

3.1.4   Timeframe for measuring outcomes
The majority of the studies were based on survey work at a single point in time. Ten employed some sort of longitudinal approach but these varied quite widely in the timeframe they observed.

The longest study was qualitative, conducted over six years in total, re-interviewing people annually (Orton 2010), however, to date, only the findings from the first three years of this study have been published. Most of the other studies only looked at two points in time: Elliehausen et al (2003 and 2007) looked at the credit report data at two points, three years

apart. O'Neill et al (2006) conducted two waves of interviews approximately 20-22 months apart, and Staten and Barron (2006) looked at credit data twice over a three year period. Illuminas (2008) followed up 15 people who had been interviewed two years previously. Kim et al (2003) conducted two waves of a survey 18 months apart.

The remainder looked at considerably shorter timeframes, with one re-contacting respondents six months after they had received advice and again six months later (Williams and Sansom, 2007). Collard et al (2012) carried out follow-up interviews six to eight weeks after people had received advice, and followed up a sub-sample qualitatively three to four months after the initial advice session. Pleasance and Balmer (2007) followed up participants after 20 weeks. Smith and Patel (2008) used a form of longitudinal analysis, by looking at all management data from money advice outreach projects collected over a period of one year, which therefore tracked individual's data over this period.

These studies gave an indication of which outcomes occurred in the short term, and which were longer term outcomes. Williams and Sansom (2007) found that most benefits seemed to occur within 6 months, although the amount owed to creditors continues to fall steadily after that. Interestingly though, while a considerable number of people, when asked six months after they had received debt advice, said that the advice had helped them to avoid enforcement action, when asked again at 12 months, the number saying this had fallen. Similarly, Optimisa (2013) found that 90 per cent of respondents reported that they had carried out at least one of the actions that they had agreed with their advisor within six months of receiving advice.

There was also a general sense that mental health outcomes occur very quickly; there seems to be an immediate sense of relief just talking about your problems, even if this relief isn't necessarily long lasting. In cases of high levels of debts, some measures to get out of debt – selling one's house, for example, or dealing with benefit claims – can take time, and therefore the ability to capture these outcomes needs to be a consideration when designing the survey. However, Orton (2010) found the majority of change in attitude and behaviour occurred in the first two years, or had occurred to the extent that it was clear which path the participant was heading down.

3.1.5 Method of follow-up in longitudinal studies
The majority of studies chose to re-survey the original participants to obtain self-reported measures of the outcomes. They used a range of methods to do this including face-to-face interviews (Orton, 2010); telephone interviews (Collard et al, 2012; Pleasance and Balmer, 2007) and self-completion surveys (Kim et al, 2003; O'Neill et al, 2005).

However, a few of the US-based studies used administrative data. Elliehausen (2003 and 2007) used credit reference agency data rather than primary collected data. Staten and Barron (2006) only collected data directly from the participants on the initial wave, then used their credit records to track their behaviour after the receipt of advice. Xiao and Wu (2008) used the debt management plan records of the debt advice agency to track the

pattern of payments made. Typically, the records from management data included information on number of credit cards or banks accounts held, revolving debt, total debt, and risk / delinquency scores.

An advantage of using administrative data is that it avoids the problem of attrition that occurs with any form of survey work. This is discussed in the next section. Data protection issues in obtaining credit report data should to be investigated before exploring this fully as an option.

3.1.6 Attrition in longitudinal studies

There is limited information on managing attrition from the longitudinal studies of debt advice, as this was not discussed in any depth (if at all) in the reports.

In terms of the *level* of attrition, this varied depending on the survey method and the timeframe covered.

Longitudinal qualitative studies tended to have a high retentions rate. Orton (2010) retained 53 of the original sample of 60 by year 3, almost certainly because the sample was small and it was easy to keep in touch with them, but also possibly as a result of the greater engagement from respondents that arises from building a relationship with a qualitative interviewer in face-to-face interviews. Similarly, for Dearden et al (2010), 50 of the original 57 participants remained at the end of the 12 months. The study carried out by Collard et al (2012) and based on telephone interviews also had a good retention rate (82 per cent), however, the sample was small (149) and the time frame was only two months. Illuminas do not give a rention rate.

Other studies experienced much higher rates of attrition. Williams and Samson (2007), who used a self-completion survey, experienced a reduction in their sample from 172 participants in wave one to 63 at wave three 12 months later - a retention rate of just 38 per cent. Kim et al (2003) received a 20 per cent response rate from the initial self-completion surveys mailed out in wave one, but achieved a retention rate of 57 per cent in wave two, 18 months later, once undeliverable questionnaires were taken into account. Pleasance and Balmer (2007) dropped from 402 respondents in wave one to 243 by wave two 20 weeks later, attributed to high levels of house moves among this group, along with people not answering phones calls.

The issue of retention is discussed more fully in the following chapter, which draws on non-debt-based longitudinal evaluations where attrition is discussed more fully, and in the case study below.

3.1.7 Case study - random encouragement design

Pleasance and Blamer (2007) used a randomised encouragement design in an attempt to evaluate the effect of seeking debt advice via the National Debtline.

Methodology

The sample was selected from users of a non-random selection of job seekers. 3,163 people screened, 2,256 reported that they had a debt problem and 1,657 agreed to participate.

709 were screened fully, and 402 were eligible in that they were:

1. Behind on payments for a mortgage/rent, credit/store card., personal loan, hire or credit purchase, utilities, maintenance payments, tax or court fines *and* having a problem paying the money owed
2. Experiencing real difficulty managing debt

AND they had not already sought advice.

From the 402 remaining sample, 205 were randomly allocated to the treatment group. The contact details of the treatment group were passed to National Debtline, who contacted participants, via the telephone, at a time indicated by participants as being convenient. The attrition rate was higher than expected: only 119 from the intervention group and 115 from the control group took part in the 20 week follow-up interview, which meant that the planned 50 week follow-up did not go ahead. From the 20 week follow-up, only 35% of the treatment group took up the offer of advice, and 10% in the control group sought advice independently.

Points of interest for potential future encouragement design

- The criteria of over-indebtedness may not have been tight enough, as 36% of the sample said that they no longer had financial problems after 20 weeks. This, and the low take up of advice, suggests that they had recruited too many people with 'low level' problems. However, a tighter definition may have led to more of the control group seeking advice.
- The sample was not nationally representative, or even representative of job seekers. The sample was also not typical of those who would normally use National Debtline, which may help explain the low take-up. Careful thought should be given to the nature / format of the advice offered, and the manner in which it is offered.
- The level of attrition was higher than may have been expected – this should be considered in any estimates or assumptions used in future evaluations. To minimize attrition, up to 12 attempts were made to re-contact at 20 weeks. Letters were sent to all non-responding participants' registered addresses asking for updated details. In addition, an incentive of £10 was offered if people remained in the trial. The interviews were conducted by telephone, whereas a face to face methodology may be more effective at decreasing attrition. For confidentiality reasons, National Debtline did not leave messages or say who they were until they were sure that they were speaking to the named participant.
- The aim was to have a minimum sample size of 308 (154 in each group) at 50 weeks (see p.657). This was not achieved due to the high level of attrition.

3.2 Current evidence on debt advice outcomes

Very few of the studies provided any conclusive evidence of the effect of debt advice on the outcomes that the Money Advice Service are particularly interested in exploring. As noted already, few of them were designed with any form of comparison group and most of these had not used the most robust methods to create it. Moreover, many of the measures relied on self-reporting and were often subjective. The non-longitudinal studies often asked about intended future behaviour, their satisfaction with the debt advice received, and their likelihood of using debt advice services again in future. Some used statements with a Likert scale relating to wider benefits such as stress, anxiety and relationships. Surprisingly few studies asked what respondents' debts were and how much these had changed. The studies that did explore the level of debts in detail tended to be qualitative in methodology.

However, objective measures were also used, such as outcomes of advice; whether they set up a Debt Management Plan, Debt Relief Order, Individual Voluntary Arrangement, and repayment plans generally (Elliehausen, 2003; Xiao and Wu, 2008).

As noted above some studies used administrative data to track outcomes. This included individuals' credit records held by credit reference agencies: to compare differences in ability to handle debt between those who had accessed advice and those who had not, focusing on factors such as number of accounts with balances, number of bank cards, level of total and consumer debt and the level of revolving debt (Elliehasuen, 2003 and 2007); to understand differences in outcomes between different delivery methods of advice (Staten and Barron, 2006); or to track the effect of being on a Debt Management Program (O'Neil et al, 2005 and 2006). Xiao and Wu (2008) used the debt management plan payment records maintained by a debt advice agency. These approaches have the advantage of providing objective measures but are, by their very nature, limited in the outcomes they can measure.

3.2.1 Outcomes on mental health

Many of the studies (Orton 2010; Collard et al, 2012; Toynbee Hall, 2011; Debt Resolution Forum, 2013; Smith and Patel, 2008; Optimisa, 2013) found that those who had received debt advice often reported an improvement in their mental health; others reported a reduction in the amount they were worrying (Pleasance and Balmer 2007, Williams and Sansum 2007, Turley and White 2007, Day et al 2008). For some people, this came about from the mere act of talking to someone (Turley & White, 2007; Orton 2010). Another reported an improvement in mental health because they were no longer being chased by creditors (Day et al, 2008).

However, it was hard to disentangle cause and effect in the interplay between debt and mental health issues (Rahim and Arthur, 2012). Undoubtedly there are some people for whom debt is a consequence of mental health problems, and conversely there are others whose mental health issues have been caused by the stress of having unmanageable debt. Fitch et al (2009) conducted a systematic review of the relationship between debt and mental health issues, and this concluded that while there was no evidence of a causal

relationship between mental health and debt they found *"plausible evidence from longitudinal research studies that indebtedness is often subsequently followed by mental health problems"* (p.18). The review found that different debts may have different impacts on mental health, and that level of concern over debt can be as negatively impactful to a person's mental health as the actual size of the debt. Therefore, while the measures of mental health may be subjective, positively effecting mental health may be one of the key outcomes of debt advice.

3.2.2 Repayment of existing debt

One of the more objective measures that some of the studies used was the extent to which debt had been paid down. Clearly, this measure was only applicable to studies that were longitudinal, or were measuring behaviour retrospectively. However, what evidence there is did suggest that this aspect could be affected by the receipt of debt advice.

Orton (2010) found that by year three of a longitudinal qualitative project, there had been a clear reduction in levels of over-indebtedness, with the majority of people being either debt-free or having a much reduced level of debt. This study identified "better able to deal with creditors – addressing indebtedness" as one of the most useful outcome measures. A number of other studies also found that those who had received debt advice reduced their levels of debt over time. Pleasance and Balmer (2007) found that people who had received debt advice were more likely to have paid down priority debts than the comparison group. Using data held by a credit reference agency, Elliehausen et al (2003 and 2007) found that those who received debt advice had significantly higher scores on debt reduction than those who chose not to receive debt advice.

However, Pleasance and Balmer (2007) also found no significant difference in the numbers who no longer had a debt problem between those who had received advice and those who didn't, when respondents were surveyed 20 weeks after receipt of advice. Measuring the effect of advice on the long-term reduction in indebtedness should therefore be a key aim in a longitudinal study.

3.2.3 Money management

Understanding the effect of debt advice on outcomes that were related to the ability of recipients to better manage their money focussed on a number of areas; specifically whether they were better able to make ends meet, whether they were more careful at budgeting, whether they were now moderating their credit use, and whether they had increased confidence in dealing with money matters. In the majority of studies, these measures were subjective in nature.

There was some indication that debt advice had a positive effect the on the ability to make ends meet; (Stamp 2011; Collard et al, 2012; Day et al, 2008). In these cases, the ability to manage better could be attributed to extra money as a result of paying off debt or as a result of earned income (Orton, 2010). One study (Collard et al, 2012) reported that social housing tenants who had received financial advice showed a measurable increase in making

ends meet: 76 per cent had benefited financially as a result of changes to their money management (compared with 56 per cent of the comparison group). Most commonly this involved a reduction in outgoings – on average tenants who had received financial advice were £10 per week better off than they were before receiving the advice. Optimisa (2013) also found that 67 per cent of those who received debt advice felt that their situation had improved since receiving the advice. In a more objective measure, Elliehausen (2007) found that counselling had a significant positive effect on individuals' risk score.

Given the lack of longitudinal studies, few had looked at future use of credit, and whether debt advice had encouraged individuals to moderate their credit use in the future. However, YouGov (2012) found that advice seekers were more likely to remain in manageable debt than slip into unmanageable debt than those who hadn't received advice, and Williams and Samson (2007) found that 89 per cent of those who had received advice reported that it had stopped them getting into further debt, at both 6 and 12 months after receiving the advice. Counselling was found to significantly decrease the number of bank card accounts used, the frequency of use, and the levels of revolving debt (Elliehausen 2003 and 2007) compared with those who hadn't received counselling. Toynbee Hall (2011) found that those who had received advice appeared to show a better understanding of priority debts, and affordable credit than would have been expected.

There was also some evidence of better money management among those who had received advice (Collard et al, 2012; YouGov, 2012; Orton, 2010), in particular in the likelihood of setting up a budget and in keeping track of money.

3.2.4   Bankruptcy
In the study carried out by Staten and Barron (2006), the respondents who had received debt counselling and went on to enrol on a debt management plan (DMP) (62 per cent were recommended to start a DMP although only 30 per cent actually started) had a lower incidence of bankruptcy and better bankruptcy and delinquency risk scores, than those who didn't start the recommended DMP.

3.2.5   The role of advice in outcomes
Overall, the most common roles that debt advice appeared to play in affecting outcomes was through the provision of information (Orton, 2010; YouGov, 2012; Rahim and Arthur, 2012; Turley & White, 2007; Collard et al, 2012) as well as providing emotional support through listening (Turley and White, 2007; Orton, 2010). Other outcomes that were attributed to the receipt of advice by those who received it were assistance in dealing with creditors (Debt Resolution Forum, 2013; Orton, 2010; Rahim and Arthur, 2012).

3.2.6   Variation of outcomes by respondent and delivery type.
A few of the studies had explored how debt advice outcomes vary by client profile, including socio-demographic characteristics and the severity and nature of their debt problems. Elliehausen et al (2007) found that the advice helped those with the worst problems the most, but also that the advice was sometimes too simple for clients' complex situations.

Optimisa (2013) on the other hand found no significant differences in the likelihood of debt clients feeling that their situation was better than before between people who had different types of debt, a different number of debts, or who were at different stages in the debt resolution process. This study also found that there was little difference in satisfaction or outcomes by whether clients had received one-off advice or received ongoing casework support. All the advice was delivered face-to-face, which appeared to be an effective channel in terms of building understanding and rapport, but could be difficult for those who had to travel.

YouGov (2012) found that the proportion of debt advice clients moving out of unmanageable debt within 12 months was similar across all channels of advice including online, telephone and face-to-face, although advice received by telephone was considered most helpful. Staten and Barron (2006) found no statistically significant difference between bankruptcy risk score or new account delinquency risk score between telephone and face-to-face clients, although face-to-face clients were more likely to file for bankruptcy in the subsequent two years.

4 Review of the Methodologies of Non-debt Longitudinal Quantitative Studies

The review examined the research reports of fourteen non-debt longitudinal studies, of which eight were impact evaluations that included control or comparison groups. They were selected to represent a diverse range of studies in terms of their subject, size and design. In terms of size they ranged from large government-funded evaluations with samples sizes of 10,000 and above, to much smaller studies funded by the ESRC with samples sizes of just a few hundred people (Bailey et al, 2009; Williams et al, 2001; Warnes,2010). Subject areas covered employment support, education initiatives, skills, careers, money guidance, homelessness and drug use. All but one (based in the US) were UK studies.

4.1 Research Design

4.1.1 Constructing control / comparison groups

Constructing a comparison group is a key issue for an evaluation of the impacts of debt advice services. Here we focus on longitudinal studies that have used comparison groups to explore the different approaches used.

Pilot and control designs

Of the eight studies reviewed that used a comparator, six of these used a design involving pilot and control areas. These studies were evaluations of new initiatives that were being piloted in a small number of areas across the country. The studies reviewed included evaluations of:

- Education Maintenance Allowance – an incentive programme to encourage 16 year olds to stay on in education (Ashworth et al, 2001; Middleton et al, 2003);
- ONE - a new benefits and employment service (Green et al, 2003);
- StepUP - an employment support programme (Bivand et al, 2006);
- Earnings Top Up – a new social security benefit (Marsh, 2001; Smith et al, 2001; Marsh et al, 2001);
- Pathways to Work – an employment support programme for people claiming Incapacity Benefit (Bewley et al, 2007); and
- Money Guidance Pathfinder – a money guidance and information service (Kempson et al, 2010).

In these studies, control areas were selected on the basis that they matched the pilot areas in key economic and social characteristics (such as employment/unemployment rates, earnings/income, qualifications, post-16 education rates, Year 11 exam outcomes) , but did not offer the new initiative. A difference-in-differences approach was then used to compare outcomes between the pilot and control areas. For further information on a difference-in-differences approach see page 9.

This approach (comparing pilot with control areas) is not possible for an evaluation of the impacts of debt advice as service providers already exist across the UK and have been delivering debt advices for many years. The other two studies reviewed (Killeen and White, 2000; Melhuish et al, 2010; NESS Team, 2012) were interesting in that they constructed matched comparison groups that were not based on identifying matched control areas (discussed below).

Matched comparison groups

An evaluation of the impact of Adult Careers Guidance services on those who used these services, as compared to those who had not (Killeen and White, 2000), was most similar in its aims and circumstances to a study of debt advice outcomes. This evaluation constructed a broad matched comparison group through surveying the general population of employed adults and selecting from these survey respondents a sub-sample that most closely matched Adult Guidance Users (based on their: area, sex, age, qualifications and level of job satisfaction) for follow-up surveys. The evaluation used 'nearest neighbour' and 'caliper' matching methods[2]. At the analysis stage a more precise re-matching process was conducted based on their survey responses[3]. A key disadvantage of this approach seems to be the large number of people surveyed to obtain a matched comparison group: out of 5,500 people surveyed 1,986 were selected after the matching process as being suitable for follow-up.

The National Evaluation Sure Start (NESS) (Melhuish et al, 2010; NESS Team, 2012) used survey data on participants in the Millennium Birth Cohort Study (MCS) to construct a comparison group using propensity score matching. While the design of the study originally included control areas, when it was decided a few years later to double the number of Sure Start areas these control areas could no longer be used in the impact evaluation. Constructing the comparison group involved selecting areas with similar economic and demographic characteristics to Sure Start areas and then within these areas selecting and matching individual families in the Millennium Cohort Study. A number of methodological difficulties were encountered with this approach:

- the non-Sure Start areas selected for the comparison group were not as deprived as Sure Start areas, this meant that the impact analysis had to exclude the most deprived families from the Sure Start sample;
- the evaluation of NESS and the Millennium Cohort Study were conducted by different research teams and even with close liaison between the teams the equivalence in the measures used across the two studies was not maintained to the

---

[2] Nearest neighbour finds matches that are the closest or most similar; caliper methods find matches that fall within a set pre-defined distance.
[3] Re-matching variables included: age, gender, qualifications and level of job satisfaction.

desired level (only variables measured in an equivalent manner in both studies could be used for detecting Sure Start effects); and

- the timing of interviews for each of the two studies occurred two years apart, so that the Millennium Cohort Study interviewed their cohort of children at age five in 2006-07, whereas the NESS study interviewed their sample of five year olds in 2007-09. This meant that some of the differences observed between the NESS and comparison groups could be attributable to more general changes in communities and societies over the two-year time period, rather than to the impacts of Sure Start.

4.1.2 Time scales – study duration, timing of survey data collection waves

The studies reviewed varied in the length of time that they tracked respondents, from just two months up to nine years. The shortest studies were evaluations of the Money Guidance Pathfinder (two months) (Kempson et al, 2010) and Financial Training for Adults 50+ (six months) (DePallo, M., 2014). These both measured improvements in financial capability, financial behaviours and financial well-being. For these type of initiatives it could be expected that impacts would be observed immediately or shortly after the provision of information / training. The evaluation of the ONE employment and benefits service also tracked benefit claimants for around six months (Green et al, 2003).

The longest studies were those assessing long-term impacts on child development – National Evaluation of Sure Start (seven years) (Melhuish et al, 2010; NESS Team, 2012) and on transitions into adulthood - Drug Pathways study (nine years) (Williams et al, 2001). The evaluation of Education Maintenance Allowance tracked 16 year olds over a four year period (Ashworth et al, 2001; Middleton et al, 2003) and the study of workers made redundant from the MG-Rover Longbridge car plant followed them for three years (Bailey et al., 2009).

The other studies reviewed followed respondents for between one and two years, with either two or three waves of data collection. These included:

- evaluation of the DWP's employment support Work Programme (Meager et al., (2014)
- evaluation of StepUP employment support programme (Bivand et al, 2006);
- evaluation of Earnings Top Up (Marsh, 2001; Smith et al, 2001; Marsh et al, 2001);
- evaluation of Pathways to Work (Bewley et al, 2007);
- the study on single homeless people's experiences of resettlement (Crane et al, 2011; Warnes and Crane, 2011);
- evaluation of the LSC Skills Coaching Trial (Page et al., 2007); and
- the evaluation of the Adult Careers Guidance Service (Killeen and White, 2000).

Some of these studies interviewed respondents at the start of the programme or initiative, to obtain a baseline measure, and then followed-them up at later intervals, for example at six and 18 months, or 12 and 18 months later (Ashworth et al, 2001; Middleton et al, 2003; Bivand et al, 2006; Marsh, 2001; Smith et al, 2001; Marsh et al, 2001; Bewley et al, 2007; Kempson et al, 2010; Crane et al, 2011; Warnes and Crane, 2011; and De Pallo, 2014).

Others did not conduct the first wave of data collection until several months after respondents had started the intervention. For example, the evaluation of the Work Programme (Meager et al, 2014) did not interview participants until 6-9 months after they had started on the programme, and conducted one follow-up at the end of the programme at 24 months. The evaluation of the ONE employment and benefits service (Green et al, 2003) interviewed benefit claimants five months after they initially made their benefit claim with a second interview six months later. Similarly the Adult Careers Guidance evaluation interviewed participants two to three months after they had started receiving guidance (the guidance process lasted up to three months) and again at around 12 and 24 months (Killeen and White, 2000). This was the only study that reported why these data collection timings were chosen: to collect data on both short-term (one year) and medium term (two year) outcomes.

4.1.3   Samples

Where evaluations compared outcomes across pilot and control area the sample frames used were existing administrative records. For example, the evaluation of the Education Maintenance Allowance (Ashworth et al, 2001; Middleton et al, 2003) drew a sample of eligible young people in both pilot and control areas from Child Benefit records. Similarly, evaluations commissioned by the Department for Work and Pensions, for example evaluations of ONE (Green et al, 2003), StepUP (Bivand et al, 2006), Earnings Top Up (Marsh, 2001; Smith et al, 2001; Marsh et al, 2001) and Pathways to Work (Bewley et al, 2007), were able to draw on benefit records to select a sample frame of eligible participants in pilot and control areas.

In the evaluation of the Adult Careers Guidance service (Killeen and White, 2000) the comparison group was recruited from a quota survey of employed adults. Recruitment ended when a target of 5,500 employed adults who were willing to take part in follow-up research had been reached. From these a matched comparison group of 1,986 people was selected for follow-up. This same approach was used in the Money Guidance Pathfinder evaluation (Kempson et al, 2010) where a comparison group of people living in non-Pathfinder areas who had sought information and guidance in the previous month, on the same subjects as the Money Guidance users, but from other providers, were identified via a quota survey.

None of these studies reported that they used particular strategies to target hard-to-reach groups. The only issue raised was that the recruitment of programme participants could take a long time where studies were waiting for new clients to be registered. The study on the resettlement of homeless people (Crane et al, 2011; Warnes and Crane, 2011) took 15 months to recruit 400 people; longer than expected. An issue was that the project Link Workers (part-funded by the research grant) responsible for recruiting clients into the study were not always passed on information about clients due to be resettled (and therefore eligible to take part) by staff within the organisation. With the adult careers guidance study (Killeen and White, 2000), it took six months to recruit a sample of 1,600 participants. Due

to pressures on staff delivering the careers guidance information about new clients was not immediately passed onto the research team who were conducting the recruitment process.

Survey modes / methods of data collection

The majority of studies reviewed collected interview data through face-to-face or telephone interviews, or using a combination of the two. There was relatively little information reported on the reasons why different modes of data collection were chosen. In some studies where interviews were long (90 minutes or more) and were less structured (i.e. included many open-ended questions), or involved administering tests, face-to-face methods were used for all data collection waves. Where a combination of modes were used it appeared that telephone was used for shorter interview waves and face-to-face used for longer interview waves.

4.1.4   Use of administrative data

Some of the studies reviewed used administrative or management information data in their design. The most extensive example of this was the evaluation of the Department for Work and Pension's Earnings Top Up pilot (an employment support programme to improve employment levels among paid workers) (Marsh, 2001; Smith et al, 2001; Marsh et al, 2001). Analysis of Employment Service job vacancy records (the number of vacancies and length of time to fill these) and unemployment benefit claim records (moves onto and off benefits) were used to examine differences in the pilot and control areas to help determine the impact of the programme on employment rates. The Pathways to Work evaluation (a DWP employment support programme for Incapacity Benefit claimants) similarly used benefit claim administrative data to compare differences in employment rates between individuals across pilot and control areas (Bewley et al, 2007).

Using management information data is a useful way to collect background data on programme participants that can reduce the amount of questions that needs to be collected through surveys. The adult Skills Coaching evaluation, that targeted out of work adults and funded by the Learning and Skills Council (Page, 2007), used programme administrative data to collect background information on programme participants and to compare the characteristics of participants with the general out of work population obtained from benefit record data. An evaluation of the Money Guidance Pathfinder (Kempson et al. 2010) used the data collected by service providers to obtain information about participants' use of the service as well as their background characteristics. Website management information data was also collected as part of the evaluation to provide an estimate of the number of people nationally using the website and statistics on the pages visited and the length of the visit.

4.2   Outcome Measures

4.2.1   Factual measures

Across the studies reviewed a range of factual information was collected from respondents including:

- employment / unemployment rates, length of time in employment / unemployment, type of employment, number of hours worked;
- job search activities;
- income / earnings / wage levels, changes in income / earnings, sources of income, benefit receipt;
- participation in education or training and qualifications obtained;
- expenditure patterns / changes, consumption e.g. of alcohol and drugs;
- criminal convictions;
- level of debts / debt amounts for different types of debts (e.g. rent arrears, utility bills, overdrafts, credit cards, bank loans, car loans); being contacted by a credit collector, use of payday loans; and
- saving amounts, changes in saving amounts.

Where financial data was collected such as earnings, expenditure and debt levels (Bailey et al., 2009; Bewley et al, 2007; Bivand et al, 2006; Crane et al, 2011; De Pallo, 2014; Killeen and White, 2000; Marsh, 2001; Middleton et al, 2003; Williams et al, 2001), none of the studies reported on approaches they took to ensure the accuracy of this data.

The evaluation of the Money Guidance Pathfinder (Kempson et al, 2010) asked respondents who had used the service to get information on money management and those in the comparison group who had sought information on this topic elsewhere, whether or not they had undertaken a range of actions, such as: contacted creditors, reduced spending, started to save, sought debt advice, drawn up a budget, changed financial products.

The evaluation of Sure Start (Melhuish et al, 2010; NESS Team, 2012) conducted assessments of children to collect information on their development, including numeracy and literacy, as well as collecting information on health measures such as BMI (body mass index).

As discussed above, some evaluations used administrative data to collect objective information on employment rates, benefit receipt and job vacancies.

4.2.2   Subjective measures
Across the studies reviewed more subjective measures that were used included:

- satisfaction e.g. with job, housing, service received;
- awareness levels;
- attitudes to work, drug use;
- perceptions of whether barriers to work had been reduced;
- future expectations e.g. of work;
- impacts on health, general well-being and self-reported health status;
- extent of worries and anxiety about money / finances and subjective financial well-being;
- confidence e.g. in meeting financial goals; and

- motivations.

4.2.3   Use of indexes of financial well-being or behaviours

In relation to financial well-being and behaviours some studies had created indexes to measure outcomes.

The evaluation of the Money Guidance Pathfinder (Kempson et al, 2010) used measures (developed from the UK Baseline Survey of Financial Capability[4]) to measure changes in respondents' financial capability before and after receiving the guidance.

The DWP evaluation of Earnings Top Up created a financial hardship index comprising of seven measures, with one point scored for each of the following (Marsh et al, 2001):

- The respondent said 'yes' to the question: 'Is there anything you really need to buy at the moment but which you just cannot find the money for?'
- The respondent had fallen behind with at least one regular bill such as utilities, Council Tax, Water Charges, telephone, TV rental.
- The respondent had fallen behind on the repayments schedule for a loan or overdraft.
- The respondent said they worried about money 'quite often' or 'almost all of the time'.
- The respondent said that they 'don't manage very well', 'have some financial difficulties' or are 'in deep financial trouble'.
- The respondent said that over the past two years they have had debts that they have found hard to repay and that this situation occurred 'often' or 'almost all the time'.
- The respondent had fallen behind with rent or mortgage payments.
- The index for scores on each of the seven measures was categorised as follows: Score 0 – no hardship; Score 1 – little hardship; Score 2 – some hardship; and Score 3 or more – severe hardship.

The evaluation of the Impact of Finances 50+ Training Classes (DePallo, 2014) used a 15 item Financial Management Behaviour Scale to measure changes in scores before and after the training.

---

[4] The Financial Capability index covers five domains: making ends meet, keeping track, planning ahead, choosing products and staying informed and scores individuals from 0 to 10 on each domain.
Atkinson A., McKay, S., Kempson, E. and Collard, S. (2006) 'Financial capability in the UK: results of a baseline survey' Consumer research 47. London: Financial Services Authority. Further developmental work by Adele Atkinson in 2009 (unpublished) led to the design of a smaller set of questions that enable more rapid measurement of the same concepts.

Table 4.1   Financial Management Behaviour Scale (Source: DePallo, 2014)

	Please indicate how often you have engaged in the following activities in the past 3 months on a scale of 1 5: 1= never, 2 seldom, 3 = sometimes, 4 = often, 5 always.
1	Comparison shopped when purchasing a product or service
2	Paid all your bills on time
3	Kept a written or electronic record of your monthly expenses
4	Stayed within your budget or spending plan
5	Paid off credit card balance in full each month
6	Maxed out the limit on one or more credit cards
7	Made only minimum payments on a loan
8	Began or maintained an emergency savings fund
9	Saved money from every paycheck
10	Saved for a long term goal such as a car, education, home etc.
11	Contributed money to a retirement account
12	Bought bonds, stocks, or mutual funds
13	Maintained or purchased an adequate health insurance policy
14	Maintained or purchased adequate property insurance like auto or homeowners insurance
15	Maintained or purchased adequate life insurance

4.2.4   Approaches for dealing with non-response and missing values

The studies reviewed provided very little information on their approaches for dealing with non-response and missing values. Approaches that were used included:

- statistical analysis to identify any non-response bias in follow-up samples and the weighting of data so that results were representative of the whole sample (Smith et al, 2001; Marsh et al, 2001);
- the use of imputation to estimate scores for item non-response (Melhuish et al,
- 2010); and
- one study based their findings only on respondents who completed all three survey waves (USA study on Financial Training Adults aged 50+) (DePallo, 2014).

On this issue, the methodological literature discusses two approaches that are typically used (Mostafa and Wiggins, 2014): "*First, weights in longitudinal surveys are constructed to adjust or re-balance the distributions of the responders so that the relative importance of*

each cohort member's characteristic in any particular sweep is reweighted according to the importance of the characteristics of those who dropped out. In other words, if the survey is losing men over time, then men will be given higher weights than women" (Mostafa and Wiggins, 2014, p.5). Second, is random multiple imputation[5]. Multiple imputation has advantages over using weights in that it can address both item and unit non-response and can be designed to meet the specific needs of the survey data.

4.3   Maintaining Longitudinal Samples

4.3.1   Survey response rates

Reports of survey response rates achieved by the studies reviewed for follow-up interview waves (i.e. after the first interview) ranged between 47 and 97 per cent, however taking out the exceptionally high and exceptionally low figures, most were between 57 per cent and 78 per cent. In some studies the response rate (calculated as a percentage of respondents who answered the previous wave) was fairly consistent across waves, whilst in others the response rate was lower in later waves. Some studies that used different interviewing modes reported lower response rates for waves that were conducted by telephone and by post (this is discussed further in Chapter 5). Attrition between waves occurred because: participants no longer wanted to take part; they were unable to be contacted during the fieldwork period (after several attempts to contact them); they were not traceable - telephone numbers no longer worked or participants were no longer at the same address; they were unable to take part due to illness; or interview appointments were broken.

It is important to note that attrition between waves has a cumulative impact on the sample size. For example, with a sample of 1,000 interviews at Wave 1 a retention rate of 70 per cent at each wave would equate to a sample of 700 completed interviews at Wave 2 and 490 at Wave 3.

In addition to attrition between interview waves, some of the studies reviewed experienced high levels of sample attrition between drawing the sample and completing the first wave of interviews. Attrition at this stage occurred because: people opted-out (where opt-out processes were conducted); people in the sample frame when contacted were no longer eligible to take part, for example, they had moved out of the pilot area, they were no longer unemployed; the contact details for sample frame members were incorrect / out of date; or people could not be contacted. Incorrect contact details was a particular problem reported in the Pathways to Work evaluation (Bewley et al, 2007), as shown in the table below.

---

[5] Multiple imputation uses statistical analysis to replace missing data with substituted values.

Table 4.2   Pathways to Work: Numbers sampled and interviewed (Source: Bewley et al, 2007)

	Pre Pathways sample	Post Pathways sample
Issued sample:	10,487	10,883
of which, responded to wave 1 interview	5,884	6,273
of which, replied to final outcome interview	2,760	3,899
of which, live in 'right' area	2,517	3,267

Of the studies that reported on the attrition rate between drawing the sample frame and completing the first interview (these were the large government funded studies: Ashworth et al, 2001; Middleton et al, 2003; Bivand et al. 2006; Bewley et al, 2007), response rates at wave 1 ranged between 45 and 63 per cent of the total sample frame.

4.3.2   Minimising attrition

Only a few studies reported in detail on their strategies for minimising attrition. Of the approaches to maximise response rates the most commonly used was the provision of incentive payments for each completed interview.

The Drug Pathway into Young Adulthood study (Williams et al, 2001) that succeeded in following up 60 per cent of participants nine years later attributed this to the strong attachment and level of trust participants felt towards the study. That it was a local study, based in Manchester with the research team based at Manchester University, may also have contributed to this.

The evaluation of the StepUP employment programme (Bivand et al, 2006) sent participants letters around a week before the start of each tranche of fieldwork containing the researchers' contact details should participants want any further information or want to arrange an interview appointment time, or want to opt out. A change of details form (with a freepost return envelope) was also included that asked participants to confirm their address and telephone number, or asked new occupants to pass on any forwarding details.

A study on the resettlement of homeless people (Crane et al, 2011; Warnes and Crane, 2011) a highly difficult to engage group, designed a very comprehensive tracking and keeping in touch strategy that included:

- collection of comprehensive contact information for the participant and their social contacts that were checked and updated at each interview wave and permission sought to contact key-workers, relatives and friends;

- change of address / change of telephone number cards given to respondents to return in Freepost envelopes;
- the sending of Christmas cards each year;
- tracking exercises conducted prior to follow-up interview waves to establish participants' whereabouts through telephone calls, house calls, letters and inquiries at homeless people projects;
- flexible interview timings to fit in with respondents; and
- the use of the same interviewers across waves to build trust and rapport (relevant when conducting face-to-face interviews).

Overall, the tracking and retention strategy worked well (78 per cent of wave 1 respondents also completed a wave 3 interview), but this required considerable time and effort.

5 Design of longitudinal study of debt advice outcomes

There are a number of issues that need to be addressed in designing a quantitative study to measure the impact of debt advice on the people who seek it. First a decision needs to be taken on which method of securing a control or comparison group should be used and what are the possibilities for selecting a sample? Secondly, what is to be included within the definition of debt advice? Thirdly, how should the tracking be done? Fourthly, over what timeframe should the tracking take place and how many times should the outcomes be assessed? And finally, which are the most appropriate impact measures to be tracked?

In answering these questions we draw on earlier studies discussed in Chapters 3 and 4, referring particularly to the randomised encouragement impact study of debt advice undertaken by Pleasance and Balmer (2007) (see 3.1.7 for case study) as it is the one most directly comparable to the proposed study.

5.1 What form of 'debt advice' is to be assessed?

The most important issue here is that debt advice comes in many forms, both in terms of the method of delivery and the nature of the advice given.

Taking the delivery method first, debt advice is delivered face-to-face, by telephone and via websites. It can involve a self-help approach or direct assistance by a debt adviser.

Most commonly debt advice results in an informal plan to repay creditors – either arranged by an adviser or done by the user him/herself. But it can also result in the setting up of formal debt solutions such as an Individual Voluntary Arrangement (IVA), advice to apply for bankruptcy or a Debt Relief Order (DRO), or an adviser applying for these on the user's behalf. These are all different interventions and the nature of the intervention could well influence the outcomes for the person being advised or assisted. Moreover, some agencies will go beyond this and, for example, either advise or assist users: to maximise their income or minimise their outgoings; to consolidate debts; or provide advice and assistance with budgeting. This assistance may be given in addition to, or instead of assistance with repaying creditors or applying for bankruptcy. Again these will influence the outcomes.

Then there are considerations relating to the diversity of debt advice providers, who provide advice through different channels (telephone, face-to-face, or interactively over the internet), and whose clients differ quite markedly in their socio-economic characteristics (most significantly income levels and linked to this economic activity status, debt levels and types of creditor to whom money is owed).

The greater the diversity of advice encompassed the less clear cut the results will be, unless the sample sizes are large and each intervention has a sample that is large enough for meaningful analysis and a bespoke comparison group (see section 5.7.2). The way round this, *and the one we would recommend, would be to focus on the most common type of intervention(s).*

5.2 Which method of securing a control or comparison group should be used?

As noted in Chapter 2, the gold standard of impact evaluations is a **randomised control trial** (RCT). Not only does it give the most robust comparison group, but it removes much of the practical difficulty of recruiting the comparison group. However, it is difficult to see how this could ethically be applied to a study of debt advice. Discussion of the subject of ethics in RCTs has generally focussed on medical research. Resnik (2008), for example, concludes that "Withholding interventions from research subjects can be ethical, provided that it does not lead to exploitation of individuals or groups". While this could be used as an argument for conducting a debt advice RCT, it could also be argued that there is no 'equipose' (Nardini, 2014) in the provision of debt advice; there is no debate among practitioners that some people would benefit from *not* receiving debt advice.

To conduct a proper RCT for debt advice, it would require everyone who approaches a debt advice agency for assistance to be intercepted and half of them to be turned away or given an appointment sometime in the future when the evaluation is complete. It is difficult to see how this could be considered ethical or that it would be acceptable to advice agencies participating in the study. The consequences of being turned away and not receiving help may be quite catastrophic for the individual and their household.

5.2.1 Randomised encouragement: the recommended approach

A more promising alternative is a **randomised encouragement approach**, where a group of people who are judged likely to benefit from debt advice is randomly assigned to either a group that is encouraged to use a debt advice agency or to one that is not. This approach has been used by Pleasance and Balmer (2007) *and is the one that we would recommend*. Even so, some might consider it unethical to identify people who need assistance and then not tell them about it.

There are also some practical problems to be addressed. Foremost of these is how to define the people who would benefit from debt advice. People contact debt advice agencies at various stages in their 'debt journey', so it would be necessary to discuss the best definition to use with a group of debt advisers drawn from across a diverse range of agencies, or to explore the option of analysing management data to understand the characteristics of the clients. Pleasance and Balmer (2007) used an objective measure – people reporting that they were behind with payments on one or more of their household commitments and having a problem repaying the money owed – and a subjective one – the extent to which people said they were experiencing difficulty managing debt. In practice this definition seemed too loose, as only 35 per cent of the treatment group took up the offer of debt advice when contacted by National Debtline and, after 20 weeks, 36 per cent of their sample said that they no longer faced financial difficulties. In the proposed study *we would*

recommend that the screening questions should reflect the circumstances of people who currently seek debt advice more closely.

Having defined the group of interest the next stage would be to find a sufficiently large sample of people. Here there are a number of options. It could, in theory, be done using either an Omnibus or a bespoke survey. The incidence of the people who would qualify is relatively low so that the costs of a bespoke survey are likely to be prohibitive; an Omnibus, however, remains an option that could be considered.

A third option is to use one of the panels that are maintained by most commercial fieldwork companies for survey purposes and to screen them for inclusion in the sample. The potential problem here is the representativeness of the people in the panel. Nevertheless, it is an option that could be considered alongside an Omnibus screening.

A fourth possibility is to 'piggy-back' off another survey that has already identified people in financial difficulty and undertake a more detailed screening against the matching criteria. There are two surveys that could, potentially, be used in this way: *Understanding Society* and the *Wealth and Assets Survey*. The potential of using existing panel surveys was considered as an option when exploring the feasibility of setting up a panel on Universal Credit (Laurie et al, 2015). However, they concluded that there was unlikely to be enough space to include the whole set of questions needed within the existing survey questionnaire, and also noted that long time frame for data collection and data release may be problematic. This would be likely to rule it out for the proposed study of debt advice too

A fifth option is to use the records of creditors or credit reference agencies to identify a group of people who are recorded as being in default on payments. Here data privacy issues are likely to be encountered.

It is, therefore, recommended that screening using an Omnibus survey or a pre-existing panel should be considered.

The next consideration is the choice of debt advice agencies that the treatment group is encouraged to use and the form that the encouragement would take. The Pleasance and Balmer study (2007) only referred to National Debtline; but the greater the variety of services referred to the larger the number of people who would have to be recruited to the treatment group to give meaningful results on the impact of debt advice. Deciding how best to encourage people to use debt advice is tricky, cash incentives would clearly be undesirable as they would potentially distort the impacts measured. It would also be important for agencies that the treatment group were encouraged to use to offer priority appointments[6]. This may restrict the agencies that would be willing or able to participate in the research and raises the risk that those who do participate may not be representative of advice agencies as a whole. The study may also therefore need to be limited to a number of

---

[6] Potentially could use a 'hot key' system that has been trialled by lenders dealing with customers in mortgage arrears, which allowed agents to transfer borrowers directly to a third-party debt advice agency.

locations. Pleasance and Balmer (2007) overcame these difficulties by passing the contact details of people selected for the treatment group directly to National Debtline who then made contact with the potential users. *We would recommend adopting this approach.*

Another important point that needs to be taken into account is that not all of those encouraged to contact a debt advice agency will actually do so. And conversely people assigned to the control group may in fact seek advice of their own accord. Even with National Debtine taking the initiative to contact the potential user, only 35 per cent of those selected for the treatment groups took up the offer of advice. While 10 per cent of the control group had sought advice of their own accord (Pleasance and Balmer ,2007). The longer the period of tracking the more likely the control group is to have sought advice. Early findings from the latter stages of the qualitative study of debt advice by Orton et al (2010), found that all of a comparison group eventually sought advice to help them deal with their mortgage arrears.

Consequently, it is likely that the study would end up with four groups: those in the treatment groups who received advice, those in the treatment group who didn't, those in the control group who independently sought advice and those who didn't. It is important to keep in mind that if an encouragement design is used, the causal impact of the program can't be evaluated by comparing the outcomes of those who participated in the program with those who did not participate. The people who choose to participate in the program may be different from the people who choose not to participate, and these differences can drive the differences in outcomes. One option for estimating the causal impact of the debt advice is to compare the outcomes of the group that was encouraged to use debt advice with the group that was not regardless of whether or not they sought advice. This comparison is known as the 'intention-to-treat' estimate because it compares the outcomes of those who were intended to be treated with those who were not. However, in the proposed study of debt advice the Money Advice Service is interested in the impact of that advice on the participants who actually receive it and the 'intention-to-treat' estimate is not the appropriate measure for this purpose. There are statistical methods to deal with this, as discussed more fully in the appendix. Put simply this means distinguishing within the treatment group those who would have sought advice even if they had not been encouraged to do so, from those who only sought advice because they were encouraged. The impact of debt advice is then estimated by comparing the outcomes of this second group with those in the control group who did not seek advice. This has implications for sample sizes and, based on the Pleasance and Balmer (2007) study this would mean that 10 per cent of the treatment group would probably have sought advice of their own accord (assuming it was the same as the control group), leaving just 25 per cent of the treatment group who probably obtained advice only because they were contacted by National Debtline. Careful consideration will also need to be given to the design of questions to identify the people among the treatment group who would have obtained advice even if they had not been encouraged to do so.

5.2.2 0ther approaches that would be less feasible and robust

Two other possibilities for the proposed study of debt advice are **propensity score matching (PSM)** and **broad matched comparison group**, which would need to be accompanied by a **difference in differences** approach to measuring impact. In both cases, a sample of debt advice clients would be identified from the records of partner debt agencies as soon as possible once they have had their initial interview. These same records (perhaps supplemented by a short interview) would be used to collect information about key characteristics of debt advice clients to which the comparison group would be matched and the selection of these characteristics is key in determining the robustness of the sample. It is, however, very likely that key attributes such as desire to resolve the debt situation cannot be captured, which would rule out both robust propensity score matching and a robust matched comparison group.

Finally, in theory, a **full-population longitudinal study** could be carried out, but this has been ruled out on cost grounds because, at the baseline, the incidence of people who experience debt problems of the severity that causes them to seek debt advice is likely to be well under 5 per cent (Kempson et al, 2004).

5.2.3 Comparing the impact of different forms of debt advice: an alternative approach

There is one further possibility if, ultimately, it proves too difficult or expensive to recruit and track a comparison group, and that is to compare the impact of different types of debt advice – say a self-help approach compared with direct intervention by a debt adviser. This could be done by random assignment, passing the details of the potential user to one of the participating debt advice agencies so that they can attempt to make contact (as was done in the Pleasance and Balmer study(2007)). While this will not provide evidence of the impact of debt advice relative to receiving none, it would allow an assessment of the relative impacts of different methods of delivery. Users would be recruited and the impacts assessed in similar ways to those outlined for the randomised encouragement approach described above.

5.3 How should the tracking be done?

There are two main possibilities for tracking outcomes: surveys and use of administrative data. Both have been used in previous studies of debt advice.

Administrative data can provide the most objective measures of outcomes and by using it, the problems associated with longitudinal studies are side-stepped. In reality, though, administrative data can probably only be used to monitor payments to creditors or people's credit records. This, in turn raises practical issues relating to gaining access to such data. If a way can be found to overcome these then it offers the best quality data and ***would be recommended.***

For this reason, there will need to be at least some (if not total) reliance on outcome information collected through surveys. Compared with using administrative data, the information collected in this way will be subjective self-reporting, the reliability of which is difficult to assess. It also leads to the challenges associated with longitudinal surveys that are discussed in the next sections.

5.4 Survey mode

The first interview wave would need to be conducted face-to-face using CAPI (Computer Assisted Personal Interviewing) for the following reasons:

- Firstly because the interview will need to be longer to capture background information about respondents' circumstances.
- Secondly to encourage and motivate participants to take part in both wave 1 and subsequent waves, explaining the importance of the longitudinal element and what their future involvement will entail.
- Thirdly to collect sufficient contact details for tracking participants and to explain what they should do if they move address or change telephone number.

Whilst the most expensive survey method, face-to-face interviews have a higher response rate (Laurie et al, 2015; ONS, 2012) which is important at this first stage.

Subsequent interviews, assuming that the questionnaire length is relatively short (around 20-30 minutes) could be conducted using an online survey (online surveys can also be completed on a mobile phone) or by telephone interviews, or possibly a combination of both (discussed below). This would have significant cost savings over face-to-face interviewing. However, there are disadvantages to using telephone or online surveys compared to CAPI, these are: lower response rates, mode effects and reduced data quality (ONS, 2012).

Analysis by Laurie et al. (2015) estimating panel attrition using different interview modes (for a longitudinal study of Universal Credit recipients) showed that online surveys have the lowest response rates estimated at 40-50 per cent of participants responding. Telephone interviews (Computer Assisted Telephone Interviewing) generally have higher response rates, but lower than for CAPI which can achieve response rates of up to 80 per cent. However, online survey response rates can be boosted by sending reminders to those that have not completed the online survey (by text, telephone, or email).

An important issue to bear in mind in using a mix of interview modes is measurement effects (Laurie et al, 2015). Surveys administered by an interviewer (using CAPI or CATI) can result in different responses to surveys that are self-administered using online methods or self-administered with an interviewer present. Ideally, either self-administered or interviewer-administered would be used, rather than a mix of both. If any questions are better delivered by self-completion, due possibly to the sensitivity of the subject matter,

then there is the possibility of using computer assisted self-interviewing (CASI) during wave one to mitigate any mixed mode measurement effect that may arise (Laurie et al. 2015)

In terms of data quality, face-to-face interviews produce higher quality data: respondents can be supported in answering questions that they find difficult to understand; showcards can be used to improve recall and recognition; the interviewer can develop a rapport with the respondent that can lead to them being more open about their circumstances (ONS, 2012).

It is therefore recommended that at least the initial interview, and ideally all follow-up interviews, be conducted face to face.

5.5 Over what timeframe should the tracking take place and how many times should the outcomes be assessed?

Whether administrative data or surveys are used to track outcomes, consideration needs to be given to the timeframe over which the tracking should be done, and also to how often, and at what points in the overall time frame, data should be collected on outcomes. Based on the measurement of outcomes from previous studies, we *would recommend up to four waves of interviews:*

- *one initial data collection before anyone receives any advice;*
- *one at six months after receiving debt advice (or not);*
- *ideally one at 12 months; and*
- *one at 20/24 months.*

This schedule should capture most of the outcomes that could reasonably be expected to occur. As noted in section 3, many outcomes were found to occur in the first six months, and the longitudinal studies based on credit report data suggested that longer term changes in credit behaviour may be seen around the two year mark. It should, however, be remembered that the longer the time period, the harder it becomes to attribute changes to the debt advice, and the greater the likelihood of members of the comparison group seeking debt advice on their own accord and the greater the dropout rate of respondents from the survey. This last point is covered in more detail in section 5.7 below.

5.6 Sample Attrition

In any longitudinal survey some people will either not be contactable or will decline to take part at each wave of interviews. This level of attrition would depend on which research design and which survey method is used. Loss of potential research participants from the sample frame during the process of completing the first interview wave is a key issue when the sample frame is drawn from a database of client records – as would be the case with using a PSM approach. The experience of other studies shows that it might be expected to lose around 50 per cent of the sample due to people opting out of the study, not being contactable, or no longer being eligible to take part. However, this issue is less applicable

when participants are recruited to the research study at the point at which they join a programme, as up to date contact details can be collected and eligibility criteria checked. But another difficulty other studies show (see Chapter 4) is that this approach to recruitment can be very time consuming as it is dependent on the level of programme take-up. Moreover, not everyone who joins the programme will want to participate in the research, and some people may still opt out when approached for an interview, or be unavailable.

Based on the evidence of other longitudinal quantitative studies (see Chapter 4) an interview response rate of 60 to 70 per cent for each follow-up interview wave could be expected, depending on the mode of interview used (as discussed above). However, the research done by Pleasance and Balmer (2007) gives some indication of the level of attrition likely with a study of people with financial difficulties. After only 20 weeks they had a retention rate of 58 per cent of the sample of people originally interviewed – putting it at the bottom of this range. This would result in a large decline in the total sample size by the end of the study. So by the third wave of follow-up interviews just 36 per cent of the original sample may be retained.

5.7.1 Minimising attrition
Here we identify best practice in minimising attrition in longitudinal studies. It draws largely on Laurie et al.'s (2015) recommendations for a Universal Credit panel, but also on other methodological literature (Calderwood, 2012; Calderwood, 2010; ONS, 2012). Best practice includes both prospective and retrospective tracking procedures:

- At wave 1 (conducted face to face) a minimum of six or more calls should be made to each address (at different times of day and different days). Only experienced interviewers should be used to recruit and interview participants.
- The use of incentive payments for each interview wave, either paid up-front (unconditional) or on completion of the interview (conditional). Unconditional incentive payments have been found to result in better response rates.
- Collection of sufficient contact details for tracking participants and checking of these at each interview wave, including email addresses (ideally more than one), telephone numbers (ideally more than one) and the names and contact details of two stable contacts e.g. friends or relatives. Permission to trace participants through social media e.g. Facebook, Twitter, LinkedIn can also be sought. The Millennium Cohort Study (Calderwood, 2010) also tracks sample members using publicly available records such as Post Office, electoral and phone records which are available on the internet or through specialist software and through other administrative data sources. These are additional information sources that could be used to track participants.
- The use of change-of-address cards (or reply slips) given to participants so that they can notify the survey agency of their new contact details. Incentives can be used to encourage participants to do this.

- Between wave mailings[7] to keep in touch with sample members and keep their contact details up-to-date. They remind sample members to get in touch with the survey team to confirm or update their contact information and can include the use of change-of-address cards. They also remind sample members about the survey and why their participation in it is important. Returned letters marked "not known at address", can be used to trigger tracing procedures before the fieldwork commences. Between wave mailings are sent in the few weeks prior to fieldwork commencing, but if the period between interview waves is more than six months, a between wave 'keep in touch' mailing may also be needed.
- During fieldwork interviewers should also attempt to track families who have moved through visiting last known addresses and asking neighbours.
- All communication with sample members should include details of how they can contact the survey team – Freepost address, Freephone number or email address. Written communications need to be easy to read (using simple language), taking into consideration the language needs of those whose first language is not English and those with poor literacy.
- The creation of a project website to provide information about the survey and the research team to help engage participants and earn their trust.

5.7.2 Likely sample sizes needed

The likely sample sizes needed will be determined not just by sampling issues but also by the level of disaggregation required in the analysis. Debt advice is not monolithic and the analysis will almost certainly want to take this into account. So if a range of debt advices are covered it will be important, at the very least, to separate out self-help from assisted clients. It may also be important to consider advice delivered face-to-face, by telephone or and on-line separately.

For simplicity's sake the sample calculations in the table below are based on a single provider (as in Pleasance and Balmer (2007) who only studied users of National Debtline. It gives the numbers required should a similar exercise be attempted again, taking into account their experience of attrition and also the need to take account of the people who would naturally seek advice (as discussed above). For ease of calculation these, like the original calculations made by Pleasance and Balmer, are based on requiring a final sample size of 308 (154 in each of the control and treatment groups). This was calculated as the minimum sample size needed to detect a doubling in the likelihood of the treatment group no longer facing debt problems after 50 weeks compared with the control group. It was calculated on the basis of a two-group continuity correct chi squared test with a 0.05 two-sided significance level where (beta)=2.

---

[7] The Wealth and Assets Survey conducts their keeping in touch via telephone to collect information about household members, to confirm or update contact details, providing an early opportunity to identify movers (ONS, 2012).

It should, however, be borne in mind that this gives an absolute minimum initial sample and is illustrative only. Even so, it suggests that around 22,500 people would need to be screened in order to give final samples of 154 people in each of the treatment and control groups after 4 waves of interviews (an initial interview and three follow-ups). It does not adjust for the use of a tighter definition of debt problems than the one used in the Pleasance and Balmer (2007) study. This would reduce the number of people potentially eligible for inclusion in the study this would counterbalanced by an increased proportion of them seeking advice - both of their own accord and when encouraged to do so.

		Group	
	Min. N needed	Treatment	Control
Wave 4: 20-24 months	308	154	154
Response rate / retention 58%			
Wave 3: 12 months	531	266	266
Response rate / retention 58%			
Wave 2: 6 months	916	458	458
Response rate / retention 58%			
Wave 1: 0 months	1,579	789	789
Total sample needed before advice take-up adjustment	4,034	3157	877
Proportion of treatment group encouraged to seek advice 25%		3,157	
Proportion control group not seeking advice = 90%			877
Response rate from screening and recruitment = 18%			
Start sample for screening	22,412	17,540	4,872

5.7   Which are the most appropriate impact measures to capture?

The most commonly measured impacts in the studies reviewed were: improved money management, the extent the client felt that their financial situation was under control, the course or courses of action they had followed as a result of the advice, and whether the receipt of debt advice had improved their mental health. While these are subjective measures, all of these are possible to measure over time, and in our view, are important measures of the outcomes of debt advice to capture.

One key measure commonly lacking in previous research (primarily due to lack of quantitative longitudinal studies) was a measure of the level of debt each participant held and changes to this over time. It is important to capture the types and amount of debt held at each wave, and the level of repayments to understand whether the advice has practically helped the recipient out of debt.

It is important at the initial wave to capture data that allows a comparison group to be made that reflects the key factors; this would include data on level of debt, length of time that debt has been a problem, socio demographic data, household income, but also a few key questions around attitudes to money, possibly derived from the Baseline Survey of Financial Capability (Atkinson et al, 2006). Key external events, particular those that relate to changes in household income such as redundancy, or divorce, should also be captured, as these often are predictors of a move into debt (Hartfree and Collard, 2015). Early findings from the final stage of the Orton longitudinal study suggest that those who were debt free, or were moving out of debt had certain characteristics in common, such as being in full time employment, free from physical and mental health problems, less likely to have young children and having confidence in dealing with financial matters.

It is important to bear in mind that the measures should be applicable to a comparison group, and should therefore be worded in a manner that doesn't assume the receipt of debt advice (e.g., asking about changes since x date, rather than since receiving advice).

Broadly speaking, all outcomes would be measured at each wave, to gain an understanding of which effects occur when, and how long-lasting any effects may be.

Below is a table that summarises the impact measures needed to measure The Money Advice Service outcomes, and which studies may have existing measures of these outcomes, or would help in constructing the new measures.

THE MONEY ADVICE SERVICE OUTCOMES			MEASUREMENT NEEDED
Recognising and framing debt problems	1	The client got the advice or information they wanted	Self reported levels of satisfaction (Collard et al, 2012; YouGov 2012)
	2	The client is aware of the possible different ways of dealing with their debt problem	List of options actions and awareness of these (Optimisa, 2013; Collard et al, 2012; YouGov, 2012)
	3	The client better understands that debtors and creditors have rights and obligations and has some idea what these are	Not currently measured, but possible to construct a measure if needed
Action to resolve debt problems	4	The client has a clear idea about the steps required to try and sort out their debt problems	Subjective measure of understanding (Optisima, 2013; Rahim et al, 2012; Turley et al, 2007)
	5	The client takes the steps they should to sort out their debt problems	Whether a particular course of action had been followed (Optimisa, 2013; YouGov, 2012)
	6	If a referral is made, the client follows it up and receives help from another organisation	Course of action taken (see above)

Money Management	7	The client increases their household income	Increase in receipt of benefits or change in work patterns (Collard et al, 2012; Orton, 2010).
	8	The client is able to plan how they use their money and sticks to the plan	To what extent money is watched, budget drawn up (Optimisa, 2013; Kim, 2003; Collard et al, 2012)
	9	The client reduces their outgoing	(see above)
	10	The client pays priority bills first before non priority (such as consumer credit)	A list of debts and which has are being paid down (Pleasance and Balmer, 2007)
Towards debt resolution	11	The client's financial situation is brought under control	Subjective: How do they feel about the manageability of their finances currently (O'Neill, 2005; Collard et al, 2012)
	12	The client knows to review their situation if their circumstances change	Not currently measured, but possible to construct a measure if needed
	13	The client's outstanding debt is reducing	Current outstanding balance of debt, assumed length of time to finish paying (Orton, 2010)
	14	The client has appropriate contact with his/her creditors	
	15	The client's wellbeing improves	Subjective, but quite key (Fitch et al 2009; O'Neill, 2005; YouGov, 2012)

5.8   Conclusions and next steps

A longitudinal study to assess the impact is feasible, using a randomised encouragement design, but it likely to be expensive. Even the simplest case of assessing the impact of one type of debt advice, is likely to require around 22,500 people to be screened to ensure that a minimum of 308 people remain in the sample after four interviews spread over 24 months. This would almost certainly mean that an Omnibus would have to be used for the screening. Potentially further research into minimising attrition could be conducted. It is also recommended that a tighter definition of financial difficulties is used in the screening than the one used by Pleasance and Balmer (2007). This will affect the sample calculations.

However, before taking this further discussions are needed within the Money Advice Service about the scope of the study and the level of analysis required. Debt advice is not a monolithic process. It includes self-help advice as well as direct assistance. It typically involves reaching informal arrangements with creditors but can also involve a range of other debt solutions, including Debt Relief Orders, Individual Voluntary Arrangements and Bankruptcy. Each debt solution included would increase the sample size and cost. It is therefore recommended that, as a minimum, there should be separate samples of self-help and assisted advice and that the focus should be on people assisted with reaching informal

arrangements with their creditors, since these form by far the largest group of users of debt advice agencies. This would increase the number of treatment groups to two and increase the initial sample required at the screening stage to at least 40,000 - although this makes no allowance for reductions in the sample to allow for people advised to follow other courses of action. A decision also needs to be taken on the most appropriate screening questions to identify people who are similar to current users of debt advice in terms of the severity of their financial problems. Once these decisions have been made detailed calculations can be made of the sample sizes and overall design and, from this, an indicative cost of the fieldwork and analysis obtained.

If sufficient finances can be found for the screening, it is feasible to measure impact and a number of potential outcomes have been suggested as impact measures. It is, however, a complex piece of sample design and analysis and would need to be undertaken by statisticians with prior experience of research of this kind.

6 Appendix

6.1 Annex: Some Comments on Survey Design and Estimation

Peter Lynn, University of Essex, March 2016

1. Survey Design

On the assumption that an RCT must be ruled out on ethical grounds, this annex comments on the two most promising design approaches: the **randomised encouragement design** (probably the preferred option) and a **full-population longitudinal study** (which may be powerful statistically, but not necessarily cost effective due to the scale required). Both of these approaches avoid the need to form a control group based only on a limited number of variables of the kind that can be obtained in a large-scale screening exercise. A control group approach would rely on unwarranted and untestable assumptions, that would ultimately cast serious doubt on the estimates of effect sizes. The inevitably large cost of a complex study of this kind would be hard to justify for a design that may not deliver robust results. The randomised encouragement design and full-population longitudinal study both overcome these limitations, though in the case of the randomised encouragement design this is only the case if the design is implemented in a thorough way which involves some complexity and has resource implications.

The randomised encouragement design relies largely, but not entirely, on randomised allocation to two groups to provide the necessary statistical control mechanism. The reason that further control is necessary is the imperfect association between the randomised treatment groups and whether or not a person seeks advice, as explained in the report. A number of statistical techniques could be used to provide this control, but the basic concept is perhaps most easily explained if we assume a matching technique of some kind. Figure 1 denotes by A, B, C, and D the four groups described in the report:

A: In the control group; sought advice

B: In the control group; did not seek advice

C: In the treatment group; sought advice

D: In the treatment group; did not seek advice

These correspond to three population subgroups:

1:      Proactive advice seekers (seek advice even without encouragement)

2:      Reluctant advice seekers (seek advice only if encouraged)

3:      Advice shunners (do not seek advice, even if encouraged)

There should be no systematic difference in outcomes between groups A1 and C1, nor between B3 and D3. Any causal effect of debt advice should manifest itself as a difference in outcomes between B2 and C2 as only the latter have sought advice. To identify this effect empirically, it is necessary to be able to distinguish between B2 and B3 and between C1 and

C2, i.e. to identify which members of the control group would have sought advice if encouraged to do so, and which members of the treatment group would not have sought advice in the absence of encouragement. This is the further control referred to earlier. It is tantamount to finding a matched group for D3 amongst B and a matched group for A1 amongst C. The conditions for this matching are good, however (compared to the conditions for relying on matching to provide the entire control group). First, the task is constrained to dividing a known group into two parts, rather than searching for a group amongst the entire population. We know that the matched group exists within the larger group and simply need to identify it. Second, the matching information is potentially much stronger. All the groups involved would be part of the longitudinal study and therefore it is possible to match not only on good predictors of outcomes (as collected in the first wave of the survey, for example) but also on the outcomes themselves (as collected at later waves).

Figure 1: Population structure and sample structure: randomised encouragement design

Group 3: Advice shunners	B3	D3
Group 2: Reluctant advice seekers	B2	C2
Group 1: Proactive advice seekers	A1	C1
	Control Group	**Treatment Group**

The corollary of this approach, of course, is that all members of both the control and treatment groups must be included at all waves of the study. The power of the statistical analysis depends on the sample sizes of groups B2 and C2. For example, if the population prevalence of group 1 is around 10% and group 3 is around 70% (broadly in line with the Pleasance and Balmer study), the causal analysis will rely on just 20% of the longitudinal study sample.

Another important point to note is that this design will directly identify the effect of advice on people in population group 2. An RCT would identify the effect on people in group 1. These effects may be different due to differences in the circumstances and characteristics of people in the two groups. To the extent that these pertinent characteristics could be measured in the longitudinal survey, the modelling of outcomes can estimate the mediating effect of these variables. This would provide a powerful tool for estimating the effect of advice on people in group 1, should this be of interest (this would provide an estimate of the impact of advice as currently provided) and for predicting the effect of advice on some or all of the people in group 3 (this would provide an estimate of the possible impact of extending advice to a larger proportion of indebted people than currently receive advice). Overall, it can therefore be seen that this is a powerful study design (and that the data collected from groups 1 and 3 are also valuable, even though not contributing to the direct estimates of effect sizes).

A full-population longitudinal study would include representative samples of groups A and B and would be able to distinguish between them. The statistical power for identifying the causal effect of advice would again come from the ability to have measured both the predictors of outcomes and the actual outcomes, in combination with the time dependency of the data. The additional strength drawn from randomised allocation in the randomised encouragement design would however be lost. However, a further advantage of the full-population longitudinal study would be the ability to fully identify the full range of dynamics involved in indebtedness. All the other designs considered, including the randomised encouragement design, are based on studying a set of people who are indebted at a particular point in time (the baseline for the study), akin to a static view of indebtedness. Such a sample will represent people in proportion to the length of time that they spend in debt. This may to some extent be thought of as appropriate, in the sense that length of time in debt may be a proxy measure of the seriousness of the debt or the need for advice, but it provides a partial picture and does not represent all people who are indebted at some point during a longer period (e.g. at any time during a particular year). A full-population longitudinal study would of course be considerably more expensive than a randomised encouragement design, as the sample would include representation of all people not in debt at baseline (possibly around 80%) in addition to those who are initially indebted. One option to reduce this expense would be to only follow-up a random subsample of those not in debt at the time of wave 1, as shown in Figure 2. Nevertheless, it seems likely that the extra cost of a full-population longitudinal study may not be justifiable, so the randomised encouragement design would seem to be the preferred option.

Figure 2: Sample structure: full-population longitudinal design

Not initially indebted			
Others initially indebted (Gps 2&3)			
Proactive advice seekers (Gp 1)			
	Wave 1	**Wave 2**	**Wave 3**

2. Practical Issues

Some important challenges with the randomised encouragement design are likely to be:

- Identifying a random sample of people currently in debt, to form the basis for random allocation to treatment and control;
- Successfully administering the treatment in a way that results in a sizeable proportion of the treatment group seeking advice (i.e. maximising the size of group 2);
- Administering the treatment in a way that is unlikely to influence participation in the study;
- Successfully recruiting and retaining a large proportion of the initial sample (i.e. minimising non-response and attrition).

The success of the study will depend on all of these challenges being adequately met. Methods to do so appear to exist, but will require some investment of time and resources.

With respect to sample identification, it could be argued that strict representation of the total population is not necessary as the basis for causal inference in the randomised encouragement design comes largely from the randomised allocation to treatment and not from randomisation in sample selection. Therefore, the use of internet panels or other non-probability sources for sample selection could be acceptable provided that the source broadly covers the population groups of interest.

One important issue is the need to ensure that allocation to treatment (and administration of the treatment) does not influence recruitment into the study. Specifically, I would recommend that administration of the treatment should not begin until after recruitment is completed: sample members in the treatment group should not be encouraged to seek debt advice until their baseline interview has been completed. This ensures that neither the awareness of advice nor the receipt of advice could affect willingness to participate in the initial interview. If such an influence were allowed, inclusion in the sample could be confounded with the outcomes of interest. Of course, such differential effects on participation (attrition) cannot be ruled out at subsequent waves, but at least by that stage the baseline data will provide considerable power for statistical adjustment to control for any effects of attrition.

As recommended in the main body of the report, the first interview wave should be conducted face-to-face. It is less clear what the most suitable and cost-effective mode of data collection would be for subsequent waves. MAS should be concerned both with the effect of mode on attrition and with the effect on measurement. It is not clear that estimates of levels of debt and other financial matters obtained in a web survey would be comparable to those obtained face-to-face. The suitability of web surveys may therefore depend on exactly what kind of information is to be sought at each wave. It is recommended that the questionnaire development stage should seek to identify possible mode effects on measurement and, as far as possible, to eliminate them through good design.

Likely levels of sample attrition will depend on design considerations that include mode of data collection, length and burden of the interview and the interval between waves and on the extent and nature of efforts made to minimise attrition. MAS clearly need a reasonable working assumption regarding non-response and attrition rates in the initial planning stages of a new study, but should then refine these once details of the features of the design become clearer and before final decisions are made about the size of the initial sample on which to base the study. A reasonable working assumption for current purposes is probably something in the region of 60% initial response, followed by 75% retention at wave 2 and 80% at wave 3.

In considering techniques for minimising attrition, targeted methods could be considered. For example, if it is possible to identify sample members at particularly high risk of attrition, these sample members could be given greater attention than others (disproportionate

resource allocation); or different persuasion messages could be used for people in different situations.

6.2 Breakdown of reports used in chapter 3

Method	Timeframe	Reports
Qualitative	Cross sectional	Day et al (2008), Finney and Davies (2011), Rahim and Arthur (2012), Turley and White (2007), Debt Resolution Forum (2013), Gillespie et al (2007), New Economics Foundation (2010),
Qualitative	Longitudinal	Orton (2010), Illuminas (2008), Dearden et al (2010)
Quantitative	Cross sectional	Xioa and Wu (2008), You Gov (2012), Citizens Advice (2014), Optimisa (2013), National Audit Office (2010),
Quantitative	Longitudinal	Elliehausen et al (2007), O'Neill et al (2006), Pleasance, and Balmer (2007), Williams and Sansom (2007), Smith and Patel (2008), Kim et al (2003), Staten and Barron (2006)
Mixed method	Cross sectional	Optimisa (2013), Toynbee Hall (2011), Stamp (2013), Dayson (2013), Williams and Sansum (2007)
Mixed method	Longitudinal	Collard et al (2012)
Literature Reviews		Hunt (2005), Evans and McAteer (2011), London Ecomnomics (2012), Fitch et al (2009), Wells et al (2010), Williams (2004)

7 Bibliography

Ashworth, K., Hardman, J., Liu, W., Maquire, S, Middleton, S. Dearden, L., Emmerson, C., Frayne, C., Goodman, A., Ichimura., H and Meghir, C. (2001) *Education Maintenance Allowance: The First Year. A quantitative evaluation*, London: Department for Education and Employment. DfEE Research Report 257

Atkinson A., McKay, S., Kempson, E. and Collard, S. (2006*) Financial capability in the UK: results of a baseline survey*, Consumer research 47. London: Financial Services Authority

Bailey, D., Chapain, C., Mahdon, M. and Fauth, R. (2009). *The Economic and Social Impact of the Demise of Rover at Longbridge: A Longitudinal and Holistic Approach to Economic Restructuring: Full Research Report,* ESRC End of Award Report, Swindon: ESRC RES-000-22-2478

Bewley, H. Dorsett, R. and Haile, G. (2007) *The impact of Pathways to Work*, London: Department for Work and Pensions. DWP Research Report 435

Bivand, P., Brooke, B., Jenkins, S. and Simmonds, D. (2006) *Evaluation of StepUP Pilot: Final report*, London: Department for Work and Pensions. DWP Research Report 337

Bryson, A., Dorsett, R and Purdon, S. (2002) *The use of propensity score matching in the evaluation of active labour market policies,* London: Department for Work and Pensions DWP Research Working Paper Number 4,

Calderwood, L. (2010) *Keeping in touch with mobile families in the UK Millennium Cohort Study*, Centre for Longitudinal Studies working paper 2010/7

Calderwood, L. (2012) *Improving between-wave mailings on longitudinal surveys: A randomised experiment on the UK Millennium Cohort Study*, Centre for Longitudinal Studies

Collard, S., Finney, A., Hayes, D. and Davies, S. (2012) *Quids in: the impact of financial skills training for social housing tenants,* Bristol: University of Bristol.

Crane, M., Warnes, T. and Coward, S. (2011) The FOR-HOME study: Moves to independent living: Single homeless people's experiences and outcomes of resettlement, Sheffield: University of Sheffield

Day, L. Collard, S. and Hay, C. (2008) *Money Advice Outreach evaluation: qualitative outcomes for clients,* Legal Services Research Centre, London: Legal Services Commission.

Debt Resolution Forum (2013) *Free-to-client advice outcomes,* Debt Resolution Forum.

Dearden, C., Goode, J., Whitfield, G. and Cox, L. (2010) *Credit and debt in low-income families*, York: Joseph Rowntree Foundation.

DePallo, M. (2014) *Impact of Finances 50+ Training Classes on Individuals' Financial Behaviours*, Washington: AARP Foundation

Elliehausen, G., Lundquist, E.C. and Staten, M.E. (2003) *The impact of credit counselling on subsequent borrower credit usage and payment behaviour,* paper presented at the 2003 Federal Reserve System Community Affairs Research Conference, Washington D.C.

Elliehausen, G., Lundquist, E.C. and Staten, M.E. (2007) '*The impact of credit counselling on subsequent borrower behaviour',* The Journal of Consumer Affairs, 41(1), pp 1-28.

Fitch, C., Hamilton, S., Basset, P. and Davey, R. (2009) *Debt and Mental Health. What Do We Know? What Should We Do?*, London: Royal College of Psychiatrists.

Green, H., Marsh, A., Connolly, H., Payne, J., Kasparova, D., Simon Kirby, S. and Riley, R. (2003) *Final Effects of ONE*, London: Department for Work and Pensions. DWP Research Report 183

Hartfree, Y. and Collard, S (2015) *'Locating credit and debt within an anti-poverty strategy for the UK',* Journal of Poverty and Social Justice, 23(3), pp 203-214.

Illuminas (2008) *Business Debtline Performance Evaluation*, Presentation, London: Illuminas

Kempson, E., Collard, S., Finney, A., Atkinson, A., Davies, S. and Hayes, D. (2010), *Money Guidance Pathfinder*, London: Financial Services Authority

Kempson, E., Finney, A. and Davies, S. (2011) *The Child Trust Fund – findings from the wave 2 evaluation,* Bristol: University of Bristol.

Kempson, E. McKay, S. and Willitts, M. (2004) *Characteristics of families in debt and the Nature of Indebtedness.* London: Department for Work and Pensions. DWP Research Report 211.

Killeen, J. and White, M. (2000) *The Impact of Careers Guidance on Adult Employed People*, London: Department for Education and Employment. DfEE Research Report 226

Kim, J., Garman, E.T. and Sorhaindo, B. (2003) *'Relationships among credit counselling clients' financial well-being, financial behaviours, financial stressor events, and health',* Financial Counselling and Planning, 14(2), pp 75-87.

Laurie, H., Buck, N., Burton, J., Parutis, V., Bryan, M., Lugtig, P. and Brewer, M. (2015) *The Feasibility of Conducting a Universal Credit Panel Survey*, DWP ad hoc research report no. 22

London Economics (2012) *Debt advice in the UK – final report for the Money Advice Service,* London: London Economics.

Lynn P, ed. 2009. Methodology of Longitudinal Surveys. Chichester, UK: John Wiley & Sons.

Marsh, A. (2001) *Earnings Top Up Evaluation: Synthesis Report*, London: Department of Social Security. DSS Research Report 135

Marsh, A., Stephenson, A., Dorsett, R. and Elias, P. (2001) *Earnings Top Up Evaluation: Effects on Low Paid Workers*, London: Department of Social Security. DSS Research Report 134

Middleton, S., Maguire, S., Ashworth, K., Legge, K. Allen, T. Perren, K., Battistin, E., Meager, N., Newton, B., Sainsbury, R., Corden, A. and Irvine, A. (2014) *Work Programme Evaluation: the participant experience report*, London: Department for Work and Pensions. DWP Research Report 892

Melhuish, E., Belsky, J. and Leyland, A. (2010), *The impact of Sure Start Local Programmes on 5 year olds and their families*, London: Department for Education and Employmnet. DfEE Research Report 067

Mostafa, T. and Wiggins, R. D. (2014) *Handling attrition and non-response in the 1970 British Cohort Study*, CLS Working Paper 2014/2, London: Centre for Longitudinal Studies.

Nardini, C. (2014) *'The ethics of clinical trials',* eCancerMedicalScience, 8, 387.

National Evaluation Sure Start (NESS) Team (2012) *Methodology Report*, available at: http://www.ness.bbk.ac.uk

Office for National Statistics (ONS) (2012) *Wealth and Assets Survey Review Report*, ONS

O'Neill, B., Sorhaindo, B., Xiao, J.J. and Garman, E.T. (2005) '*Financially distressed consumers: their financial practices, financial well-being, and health',* Financial Counselling and Planning, 16(1), pp 73-87.

O'Neill, B., Prawitz, A., Sorhaindo, B, Kim, J. and Garman, E.T. (2006) *'Changes in health, negative financial events, and financial distress/financial well-being for debt management program clients',* Financial Counselling and Planning, 17(2), pp 46-63.

Optimisa (2013) *360 degree of evaluation of Money Advice Service funded face-to-face debt advice,* prepared for the Money Advice Service, London: Optimisa Research.

Orton, M. (2010) *The long-term impact of debt advice on low income households,* The Year 3 Report, Warwick: University of Warwick.[8]

Page, R., Casebourne, J., Mason, D. and Tuohy, S. (2007) *Evaluation of the Year 2 of the Skills Coaching Trials and Skills Passports - Final report*, Coventry: Learning and Skills Council

Piesse A, Judkins D, Kalton G. 2009. *Using longitudinal surveys to evaluate interventions*. See Lynn 2009, pp. 303–16.

---

[8] See also Orton, M. (2008) *The long-term impact of debt advice on low income households,* Project Working Paper 2: The Year 1 Report, Warwick: University of Warwick and Orton, M. (2009) *The long-term impact of debt advice on low income households,* Project Working Paper 3: The Year 2 Report, Warwick: University of Warwick.

Pleasance, P. and Balmer, N.J. (2007) *'Changing fortunes: results from a randomized trial of the offer of debt advice in England and Wales',* Journal of Empirical and Legal Studies, 4(3), pp 654-673.

Rahim, N. and Arthur, S. (2012) *Experiences of debt and debt advice services in Islington – final report,* London: NatCen Social Research.

Resnik, D.B. (2008) *'Randomized control trials in environmental health research: ethical issues',* Journal of Environmental Health, 70(6), pp 28-30.

Stamp, S. (2011) *The Impact of Debt Advice as a Response to Financial Difficulties in Ireland,* Social Policy and Society, 11(1), 93-104.

Smith, A., Dorsett, R. and McKnight, A. (2001) *Earnings Top Up Evaluation: Effects on Unemployed People*, London: Department of Social Security. DSS Research Report 131

Smith, M. and Patel, A. (2008) *Money Advice Outreach evaluation: cost and effectiveness of the outreach pilots,* Research Findings 22, Legal Services Research Centre, London: Legal Services Commission.

Staten, M.E. and Barron, J.M. (2006) *Evaluating the effectiveness of credit counselling – phase one: the impact of delivery channels for credit counselling services. http://www.consumerfed.org/pdfs/Credit_Counseling_Report061206.pdf*

Toynbee Hall (2011) *Breaking the cycle – embedding financial capability within debt advice at Plymouth Citizens Advice Bureau,* Transact, Royal Bank of Scotland.

Turley, C. and White, C. (2007) *Assessing the impact of advice for people with debt problems,* Briefing Paper, London: Legal Services Research Centre.

Warnes, A. (2010) *Factors in the Outcomes of the Resettlement of Homeless People*, *ESRC End of Award Report,* Swindon: ESRC *RES-062-23-0253*

Warnes, T. and Crane, M. (2011) *The challenges of and returns from a longitudinal study of the resettlement of single homeless people*, APHA meeting Caucus on Homelessness Symposium, 1 November 2011, Washington DC

Wells, J., Leston, J. and Gostelow, M. (2010) *The Impact of Independent Debt Advice Services on the UK Credit Industry*, London: Friends Provident Foundation.

Williams, K. and Sansom, A. (2007) *Twelve months later: does advice help? The impact of debt advice – advice agency clients study,* London: Ministry of Justice.

Williams, T. (2004) *Review of research into the impact of debt advice,* London: Legal Services Research Centre.

Williams, L., Parker, H., Aldridge, J. (2001) *Drug Pathways into Young Adulthood: Follow-up of a Longitudinal Sample of Drugwise 'Post Adolescents', 1999-2000*, ESRC End of Award Report, Swindon: ESRC RES-000-23-7912

Xiao, J.J. and Wu, J. (2008) *'Completing Debt Management Plans in credit counselling: an application of the theory of planned behaviour',* Journal of Financial Counselling and Planning, 19(2), pp 29-45.

Yoong, J., Mihaly, K., Bauhoff, S., Rabinovich, L. and Hung, A. (2013) *A toolkit for the evaluation of financial capability programs in low- and middle-income countries,* Washington: International Bank for Reconstruction and Development / The World Bank.

YouGov (2012) *The effectiveness of debt advice in the UK,* Money Advice Service, London: YouGov.

University of
BRISTOL

https://pfrc.blogs.bristol.ac.uk