Survey Research Methods (2017)
Vol. 11, No. 1, pp. 93-103
doi:10.18148/srm/2017.v11i1.6734

© European Survey Research Association
ISSN 1864-3361
http://www.surveymethods.org

From standardised to targeted survey procedures for tackling non-response and attrition

Peter Lynn
University of Essex, UK

Recent decades have seen a gradual shift away from surveys in which all procedures are completely standardised towards a variety of approaches (tailored, responsive, adaptive) in which different sample members are treated differently. A particular variant of the non-standardised approach involves applying to each of a number of subgroups of the sample a targeted design feature (or set of features) that is identified in advance of field work and is not then modified during field work. Thus, the design is not "responsive". As procedures are applied to broad subgroups, they are not "tailored" either. Targeted designs may be seen as a sub-category of adaptive designs. The effectiveness of targeted designs depends partly on the richness of information available about sample members prior to field work. For this reason they have mainly been implemented on panel surveys and mainly to address non-response and attrition. This article reviews the development of targeted designs in recent years, discusses the objectives of such designs, provides a framework for consideration of these designs, and outlines ways in which targeted designs might usefully develop in the years ahead.

Keywords: Adaptive survey design; Longitudinal surveys; Mixed-mode surveys; Non-response error

1 Introduction

Currently, the predominant approach to quantitative survey research remains one of standardisation, in the sense that each sample member is subject to a standard set of procedures designed to secure the participation of the sample member and the provision of high quality data. When it comes to administering survey data collection instruments, this is for good reason. To achieve equivalence of measurement – the foundation of reliability and validity – nothing has yet been found to be superior to the standardisation of stimuli, context and so on (though the merits of a more flexible conversational approach have been discussed and evaluated by Suchman and Jordan, 1990 and Schober and Conrad, 1997). However, with respect to the many other steps in the survey process, and particularly those that are concerned with attempting to gain the co-operation of sample members, the rationale for standardisation is less clear. Nevertheless, standardisation pervades most aspects of survey design and implementation, with one exception. With regard to the initial contact with a sample member in interviewer-administered surveys, the merits of the interviewer tailoring what they say to the circumstances and concerns of the sample member were extolled by Groves, Cialdini, and Couper (1992) and Morton-Williams (1993). This is one step in the survey process where a departure from standardisation has become the norm. Most survey organisations now train their interviewers to tailor their initial introduction, whether by telephone or face-to-face. With the exception of interviewer introductions, examples of surveys that diverge from standardisation are rare, despite a wealth of survey methodological literature showing that the effects of various survey design features tend to be heterogeneous across subgroups of sample members (for example, the form and value of incentives (Singer, 2002; VanGeest, Johnson, & Welch, 2007), the length of the invitation letter (Kaplowitz, Lupi, Couper, & Thorp, 2012), and interviewer calling patterns (Bennett & Steel, 2000; Campanelli, Sturgis, & Purdon, 1997)).

The persistence of standardisation of most survey design features on most surveys may owe a lot to time and cost considerations. It is clearly cheaper and easier to design and print only one version of an advance letter, to develop and apply just one set of call scheduling rules, and so on. And with budget constraints and time pressure to get a complex survey into the field on time, it may be tempting to take the line of least resistance. But increasingly researchers have been questioning whether this approach is truly the most cost effective and have been experimenting with ways of targeting various design features to different sample subgroups. The next section of this article provides a definition of targeted survey designs and a framework within which to con-

Contact information: Peter Lynn, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex UK, CO4 3SQ email: plynn@essex.ac.uk.

sider the variety of such designs in terms of their aims, their methods, and considerations in their adoption. The following section provides an overview of types of targeted design features that have been tested and/or adopted with a view to improving response rates or sample balance and the extent to which these designs appear to have succeeded. The article concludes with a forward look to how targeted designs to tackle non-response may be extended and developed in the future.

2 Targeted Survey Designs

A targeted survey design can be defined as one in which, a) one or more design features are varied between subgroups of sample members, with the objective of beneficially affecting the relationship between survey costs and survey errors, and b) the variation(s) in design feature(s) are identified and planned in advance of the commencement of data collection and no further adaptations are made during field work (Haan & Ongena, 2014; Lynn, 2014, 2016). Unlike responsive designs (Groves & Heeringa, 2006) and the majority of adaptive designs (Wagner, 2008), all variations in design are therefore between sample units, not over time within units (there is no dynamic component). Targeted designs can therefore be thought of as a form of static adaptive design, in the terminology of Bethlehem, Cobben, and Schouten (2011).

Situations often arise in which a design feature is varied between subgroups of sample members for reasons other than affecting costs and errors. For example, different versions of a respondent communication may reflect the need to convey different information for ethical or logistical reasons, or because the survey tasks may differ between pre-identified population subgroups. Such situations share practical implementation issues with targeted designs, but have different objectives and different criteria for choosing the groups and designing the targeted feature.

Targeted designs require information about sample units in advance of survey data collection. This information is used, a) to identify subgroups to be treated differently, and b) to identify the treatment to be applied to each group. Suitable information may be available from the sampling frame, but for many surveys the sampling frame is largely uninformative. However, in the case of longitudinal surveys, once wave 1 has been completed, a wealth of information is available that can be used for targeting, including survey data and paradata from all previous waves. Targeted design practice is therefore particularly suitable for longitudinal surveys and, perhaps for that reason, has mainly been developed in that context.

A targeted design involves identifying a limited number of sample subgroups that meet three basic criteria (Lynn, 2014):

1. There should be a manageable number of groups (though the number of groups considered to be manageable will depend on the nature of the variations in treatment).
2. Each group should have defining characteristics that lend themselves to targeted treatment.
3. The groups must vary in terms of the cost of the treatment variation to be applied and/or in terms of their contribution to survey error (e.g. response propensities, if the targeting is aimed at reducing non-response error).

The first two criteria are necessary to be able to implement the targeted design, while the third is necessary for the design to be able to achieve the objective of affecting the relationship between survey costs and survey errors.

As stated above, a targeted design should have the objective of beneficially affecting the relationship between survey costs and survey errors. However, the focus of a targeted design feature will typically be just one error source or component of error. Targeted design features cannot be used to tackle coverage error or sampling error, as the targeting is applied after the sample has been selected. Targeted design features have typically been used to tackle non-response (error). The aim may be to reduce the error due to a failure to locate sample members, due to noncontacts, or due to refusals. The mechanism is often implied rather than explicit, with the stated aim being to provide better sample balance or to reduce nonlocation, noncontact or refusal rates amongst subgroups in which these rates are usually relatively high (Lynn, 2014).

The application of targeted designs to non-response is the focus of this article, but it is worth noting that there is no reason in principle why targeted design features could not be used to tackle measurement error. Indeed, such design features are already used on many surveys, though they are not typically thought of as belonging to the family of targeted or adaptive designs. Dependent interviewing (Jäckle, 2009; Mathiowetz & McGonagle, 2000) is employed by many longitudinal surveys[1] as a means of controlling measurement error in measures of change over time. Typically, this involves adapting the wording of a question to reflect the response(s) given at the previous wave(s). If the adaptation involves repeating back to the respondent the verbatim answer that they gave previously, as in Lynn and Sala (2006), then this may be thought of as tailoring rather than targeting. But other forms of dependent interviewing involve administering a limited number of versions of a question to a

[1]Examples include the Survey of Income and Program Participation (Moore, 2007), the UK Millenium Cohort Study (Londra & Calderwood, 2006), the Panel Study of Income Dynamics (Beaule et al., 2015), the Household, Income and Labour Dynamics in Australia (HILDA) Survey (Watson & Wilkins, 2012), and the German Labour Market and Social Security Panel (Eggs & Jäckle, 2015).

corresponding number of sample subgroups – a procedure that might be thought of as targeting[2]. Examples of this form of dependent interviewing include administering one of two versions of a question depending on response to a certain question in the previous wave, as in Sala, Knies, and Burton (2014), and asking certain questions only of sample members for whom the corresponding information cannot be obtained directly from the sampling frame or from a linked administrative data source, as in the EU-SILC survey (Wolff, Montaigne, & Rojas González, 2010), in which income data is collected directly from tax registers for sample members in some EU member states but must be obtained through survey questions in all states in which such registers either do not exist or cannot be used for survey purposes (Verma & Betti, 2010).

Aside from dependent interviewing, other targeted design features could be used as a tool to reduce measurement error. For example, statements designed to motivate respondents to provide thoughtful answers, in the spirit of Cannell, Miller, and Oksenberg (1981), could be varied between sample subgroups in their content, wording, or in the frequency with which they are repeated during the interview (e.g. see Al Baghal & Lynn, 2015). Designs that tackle measurement error through the use of targeted procedures such as dependent interviewing are not generally thought of as targeted or adaptive designs, but they certainly share common characteristics and some ideas from one sphere may be transferable to the other.

3 Targeted Features for Tackling Nonresponse

In targeted designs that are focussed on nonresponse, the design features that can be targeted include incentives (monetary or otherwise), field time (i.e. prioritising certain types of cases), calling patterns (i.e. varying the scheduling and callback rules between sample subgroups), the content and design of various communications with sample members (advance letters, information brochures, between-wave mailings, etc), and methods to encourage keeping in touch or notifying changes of address, telephone number or email address. Examples of targeted versions of some of these features are presented below. Each targeted design feature can be broadly categorised on three dimensions, namely the primary agent of change, the mechanism through which change is achieved, and the outcome that should be affected. The primary agent of change can be either the sample member or the interviewer. This is the person at whom the targeted feature is directly aimed. Participation is ultimately always a decision of the sample member, so in the case of interviewer-administered surveys, a feature targeted at interviewers can only be effective via their interaction with sample members. Nevertheless, the distinction between interviewers and sample members as primary agents of change may be useful to help classify targeted designs and to help understand why some targeted design features may be more successful than others. The mechanism through which sample members can be stimulated to change their participation behaviour may be either a reduction in burden or an increase in motivation. The mechanism through which interviewers can be stimulated to change their behaviour in ways that affect respondent participation may be either an increase in motivation or a reduction in barriers to effective performance. The outcome that should be affected can be location propensity, contact propensity, or co-operation propensity. The interplay of these three dimensions is summarised in Figure 1. Note that some of the mechanisms cannot be expected to influence all three of the outcomes. For example respondents can be motivated to co-operate, and they can also be motivated to provide contact details which might improve the chances of location. But it is hard to imagine a design feature that would motivate sample members to be contacted. Hence, there is no arrow in figure 1 from motivation to contact. That said, the possibility of some design feature having an effect corresponding to one of the missing arrows in figure 1 cannot be completely ruled out. Rather, the arrows should be interpreted as representing the combinations of mechanism and outcome that are likely to be the explicit objective of a targeted design feature.

Ways of improving the motivation of some respondents could include the use of targeted motivational statements in advance letters or other survey materials (Lynn, 2016), the provision of targeted feedback on survey findings (Fumagalli, Laurie, & Lynn, 2013) or other targeted materials (Cleary & Balmer, 2015), the inclusion of survey questions on topics in which the respondent is known to be interested (Oudejans & Scherpenzeel, 2012), the allocation of a particular type of interviewer (Luiten & Schouten, 2013) or the use of differential incentives. Ways in which burden can be reduced for some respondents include the provision of a more convenient mode (Al Baghal & Lynn, in progress; Luiten & Schouten, 2013) or the administration of an abridged version of the survey instrument. Interviewers can be motivated to increase effort on low propensity cases by means of differential payments or incentives (Calderwood, Carpenter, & Cleary, 2013; Peytchev, Riley, Rosen, Murphy, & Lindblad, 2010) while barriers to achieving desired outcomes for targeted cases can be reduced by targeted call scheduling (Luiten & Schouten, 2013), targeted persuasion statements (Lipps, 2012), providing more field time (Calderwood, Cleary, Flore, & Wiggins, 2012) or, conversely, capping the amount of effort to be made for cases with the highest (Beaumont, Bocci,

[2]This particular type of adaptation is similar to within-interview routing. The difference is simply that the information that determines the routing is known prior to the start of the interview. Though a subtle difference, this renders the approach consistent with the definitions of dependent interviewing provided in Jäckle (2009), Lynn and Sala (2006), and Mathiowetz and McGonagle (2000).

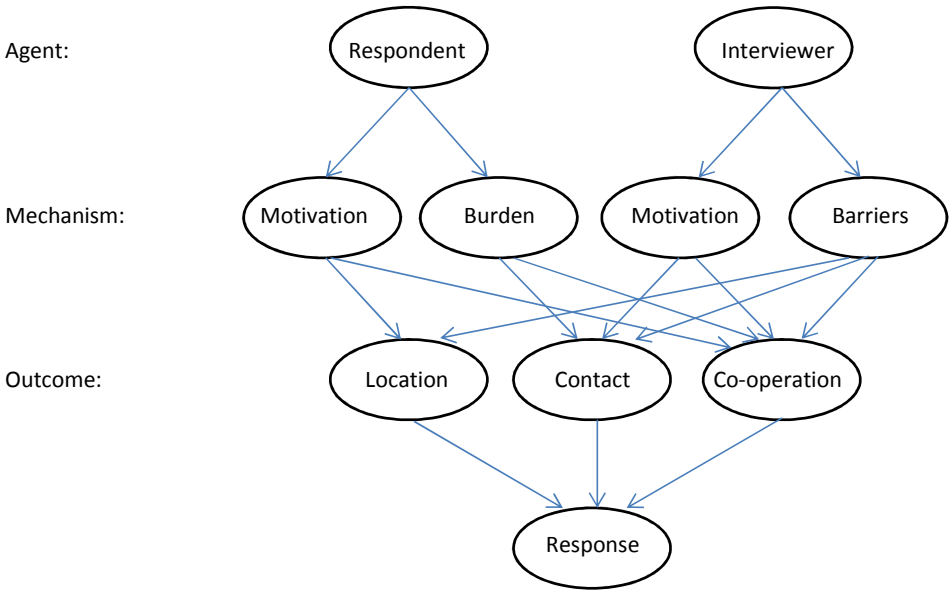

Figure 1. Dimensions of a Targeted Design Feature

& Haziza, 2014) or lowest (Johansson, Lundquist, Westling, & Durrant, 2015) predicted response propensities.

It could be argued that the potential benefits of targeted features for tackling nonresponse could equally well be achieved through post-hoc weighting adjustments. Similar arguments have been made against the idea of over-sampling strata known to have lower response rates (ESS Sampling Expert Panel, 2014). It is certainly true that the variables used to define targeted groups can also be used as weighting variables. But weighting cannot provide improvements in precision and nor does it have potential to reduce any component of nonresponse bias that arises within weighting classes rather than between classes. Targeted designs have the potential to bring improvements in both of those respects. In particular, if sample members within a targeted group who participate with a targeted feature but would not have done so in the absence of the targeted feature are systematically different from other respondents in the group, non-response bias could be reduced to a greater extent than would be possible with weighting. Indeed, there is some evidence that adaptive designs can achieve greater bias reduction than weighting alone (Schouten, Cobben, Lundquist, & Wagner, 2016). Furthermore, precision considerations are particularly important in the case of longitudinal surveys. Typically, it is not possible (for most/many analytical purposes) to increase the gross sample size once the survey has begun. Therefore the only way to maintain sufficiently large net sample sizes is to maintain high wave-on-wave response rates. This is particularly challenging for population subgroups with generally low response rates: such subgroups are strong candidates for the administration of targeted procedures.

It should also be noted that while the objective of the kinds of design features discussed here is to affect the relationship between nonresponse error and costs, some of these features could also affect other error sources. For example, a feature designed to motivate co-operation with the survey may also inadvertently motivate some respondents to answer more carefully, hence affecting measurement error. Researchers should be alert to the possibility of such unintended consequences.

4 Targeted Designs in Practice

There are several examples of targeted designs concerned with nonresponse having been implemented in practice, though many of these were experimental or trial implementations, rather than part of a full survey production process. Most of these aimed to increase co-operation rates amongst one or more subgroups with relatively low predicted co-operation rates. Some aimed to increase location or contact rates, and some aimed to increase response rates without specifying explicitly which component(s) were expected to be affected.

4.1 Location

In interviewer-administered longitudinal surveys, the probability of successfully locating a sample member at a particular wave is generally high for those who have not moved since the previous wave. Even amongst those who move, many can be easily located. The challenge (Couper & Ofstedal, 2009) is to predict which sample members are likely to move and likely to be difficult to locate if they move. Lynn (2012) built models based on data from earlier waves to predict the probability of failing to locate a sample member and then applied those models to the most recent wave in order to predict the risk of failing to locate at the next wave. Based on an estimated cost-effectiveness trade-off, the 5% of sample cases with the highest predicted probability of not being located were selected for a targeted treatment that consisted of additional between-wave mailings with requests to provide up-dated contact information. As the target group is small, the treatment could alternatively have involved a more expensive intervention such as a telephone contact, or a prepaid incentive (in the manner of McGonagle, Couper, & Schoeni, 2011). The Survey of Income and Program Participation is planning to use fieldwork prioritisation as a means to improve the location rate amongst sample cases deemed likely to move, based on previous wave data (Walejko, 2015).

Both Fumagalli et al. (2013) and Cleary and Balmer (2015) compared two alternative approaches to requesting sample members to provide address updates between waves: an "address confirmation" card to be returned by all sample members, or a "change of address" card, only to be returned by those whose details have changed since the previous wave. Both studies found that a similar proportion of address changes were reported with either method and that subsequent wave response rate did not differ between the two methods. Both conclude that the "change of address" approach should therefore be preferred, as it is less costly. Additional mailings of this kind could be sent to subgroups predicted to be at higher risk of failure to locate.

4.2 Contact

Improving the contact rate amongst the most hard-to-contact sample members was the objective of a targeted procedure reported by Calderwood et al. (2012). Indicators derived from call record data of the difficulty of making contact at prior waves were used to identify a group of cases that were then given priority at the next wave by issuing them to the field two weeks before the remainder of the sample. The effect on contact rate cannot be assessed, however, as this was not implemented experimentally. The contact rate was also a focus of attention for Luiten and Schouten (2013). During the telephone phase of a sequential mixed-mode design, different call scheduling was applied to each of four contact propensity groups. The group with the highest propensity was called mainly in the day time, and field work began later than for the other groups. Households in the lowest propensity group, on the other hand, were called in every shift (morning, afternoon, evening) every day until contact was made. The other two groups were administered intermediary call schedules. The outcome was less variation in contact rates between the four groups than in a comparable survey that served as a control group. With the targeted design, contact rates were 87.1% in the lowest contact propensity group and 95.3% in the highest contact propensity group, whereas in the comparable survey contact rates were 84.2% in the lowest contact propensity group and 96.9% in the highest contact propensity group.

Differences in the call scheduling algorithm were also the focus of Kreuter and Müller (2015). The sample for a CATI panel survey was split into 17 subgroups defined by the time window and day of the week at which they had been interviewed at the previous wave (3 time windows for each week day, plus 2 weekend windows). For each subgroup, the treatment consisted of constraining the first contact attempts to the same time window as the previous wave interview. However, compared to a control group, the treatment did not improve the contact rate, reduce the total number of attempts needed to make contact, or increase the proportion of interviews completed at first contact.

4.3 Co-operation

Reducing variation in co-operation propensities has been the aim of several targeted designs. Three studies have tested the effects of mailing targeted variants of respondent communications of different kinds in an attempt to improve respondent motivation. Fumagalli et al. (2013) produced and mailed targeted versions of a motivational results brochure between waves of a face-to-face interview survey. Three versions of the brochure were produced; two each targeted at a specific low-propensity group and a third generic version mailed to the remainder of the sample. In both of the low-propensity subgroups the targeted version improved the response rate compared to a control treatment consisting of the generic brochure. Between waves 1 and 2 of another face-to-face interview survey, Cleary and Balmer (2015) similarly sent targeted versions of a between-wave mailing, but whereas the targeted subgroups of Fumagalli et al. (2013) accounted for only 35% of the total sample, those of Cleary and Balmer (2015) accounted for 87% of their sample, and as a result their targeting produced not only improved response rates in the targeted subgroups, but also overall, from 70.6% to 75.6% (P=0.005). In a mixed-mode survey, Lynn (2016) experimented with targeted versions of the advance letter mailed to sample members who were being asked to take part in a face-to-face interview and of the invitation letter (and email) mailed to sample members who were being asked to take part in a web survey. Five targeted versions of

each letter were mailed, along with a generic version for the remainder of each sample. Compared to the use of a generic letter for the whole sample, the targeted approach improved the response rate amongst recent panel entrants in the face-to-face sample, and amongst previous wave nonrespondents in the web sample.

Allocation of interviewers to sample cases was targeted in a study by Luiten and Schouten (2013). At the telephone phase of a mixed mode survey, the best-performing telephone interviewers were allocated to respondents with the lowest predicted co-operation propensities, and *vice versa*. However, this was not successful in altering the distribution of co-operation rates, apparently because the researchers had failed to notice in advance that the differences in predicted co-operation propensities were not driven by differences in refusal rates, but rather by differences in the prevalence of language barriers and other reasons for non-cooperation (illness, absence) – for which rather different targeted interventions would be appropriate. Lipps (2012) proposed that telephone interviewers could be better equipped to deal with respondent reluctance by providing them with a targeted set of persuasive arguments, based on predicted reasons for refusal, but this intervention was not manipulated experimentally.

Two studies have attempted to raise the co-operation rates of sample members with the lowest predicted co-operation propensities by means of interviewer incentivisation. Peytchev et al. (2010) offered interviewers a larger bonus payment for each completed interview with a sample member in the low-propensity group, while Calderwood et al. (2013) implemented a slightly more nuanced version of this targeted design, with different levels of bonus depending on the predicted co-operation propensity. In both studies the targeted design failed to improve co-operation rates in the targeted group.

A number of studies have used allocation to different mode treatments to attempt to reduce the variation in response propensities by reducing respondent burden. In the initial, self-completion, phase of the Luiten and Schouten (2013) study, high co-operation propensity sample members were allocated to web mode, while low propensity sample members were sent a paper questionnaire. This allocation was intended to increase the burden for the most motivated sample members while reducing the burden for the least motivated. In this it apparently succeeded, as co-operation rates varied less between the propensity groups than in a reference survey in which all groups received an invitation to a web survey, with a mail survey option available upon request. With the targeted design, the response rate to the self-completion phase was actually slightly lower in the high propensity group (30.8%) than in the low propensity group (35.1%), whereas in the reference survey the response rate was 37.7% in the high propensity group and 18.3% in the low propensity group.

After web and telephone phases of a mixed-mode survey, Rosen et al. (2014) allocated remaining nonrespondents in the lowest quartile of the distribution of predicted response propensities to a third-phase protocol that involved face-to-face follow attempts while the remainder of the sample continued to receive only web and telephone approaches. Though Rosen et al. (2014) implemented their design as a dynamic adaptive design (the response propensities were estimated, and the allocation to third phase treatment made, only after the first two phases of fieldwork), the same design could be implemented by targeting the two alternative mode treatments to sample subgroups at the outset. Allocation to mode treatment was also employed by Al Baghal and Lynn (in progress), who allocated the lowest response propensity households at wave 8 of a panel survey to CAPI and higher response propensity households to web mode (with a CAPI follow-up of non-respondents).

Many longitudinal surveys have provided different levels of incentives to sample members, depending on their participation history. For example, the Survey of Program Dynamics provided $40 unconditional incentives to households that did not respond at the previous wave or showed reluctance previously, but no incentives to other households (Kay, Boggess, Selvavel, & McMahon, 2001). At the 2003-04 wave ("round 7") of the National Longitudinal Survey of Youth 1997, sample members who had not participated in any of the previous three waves were offered a $35 incentive, those who had not participated in either of the previous two waves were offered $30, others who were nonrespondents at the previous wave were offered $25, while all those who had responded at the previous wave were offered $20 (Bureau of Labor Statistics, undated). In the 2006 National Survey of Recent College Graduates, sample members in groups predicted to have low response rates (defined by study field) were sent a prepaid incentive, while others received no incentive. Additionally, in a second phase of fieldwork (further) incentives were offered to sample members who had not yet responded. This was done experimentally and a substantial positive effect on response rate was observed (Zukerberg, Hall, & Henly, 2007). Understanding Society, The UK Household Longitudinal Study, offers previous wave non-respondents a £20 incentive, conditional on participating, while previous wave respondents receive £10. Additionally, a further £20 can be offered when attempting to convert a refusal (Jessop & Oksala, 2014). Such tactics of offering higher incentives to motivate members of sample subgroups with the lowest co-operation propensities are consistent with methodological studies that have shown incentives to be more effective amongst sample members with low co-operation propensities (Zagorsky & Rhoton, 2008) and may improve both co-operation rates and sample balance. However, many longitudinal surveys do not provide differential incentives, or indeed any incentives (Laurie & Lynn, 2009).

5 Summary

The majority of the targeted designs that have been implemented to date, as described in the previous section, are either experimental or exploratory in nature. Few, with the exception of those involving respondent incentives, have been implemented routinely as part of survey production. However, several of the experimental studies were mounted on full production surveys. This is a useful option, particularly for regular surveys that may not have separate developmental survey infrastructure such as a test panel, though it may only be acceptable to experiment on the main survey if the risks of negative consequences of the treatments are small. Table 1 classifies the studies summarised above, following the framework introduced in section 3 above. It is noticeable that, despite the total number of studies to date being rather small, they include examples of all four of the possible agent-mechanism combinations. For each of these combinations it has been demonstrated that targeted design features can be developed and implemented. However, it is also noticeable that beneficial effects on the relationship between survey costs and survey errors have been demonstrated by all of the experimental studies in which the agent of change is the respondent but not by any of the studies in which the agent of change is the interviewer. Perhaps we should not read too much into this, given the limited number of studies and the fact that these are still early days for the implementation of targeted survey design features. But it may suggest that identifying ways of manipulating interviewer behaviour in order to bring about desired changes in field outcomes is not straightforward and is a topic deserving of further research.

Though several studies of targeted design features have been identified and summarised in this article, evidence of the effects of targeting remains limited. This may be one reason why targeted features have not been adopted as routine practice on production surveys. Most surveys have several design features that could be targeted and for each of these there are many possible ways to define subgroups to target and many different treatments that could be administered to each group. It is therefore unrealistic to expect evidence regarding each possible design. Rather, a portfolio of evidence, along with better understanding of how and why each effect comes about, can help to inform future designs. Many of the design features that could potentially be applied in a targeted way have already been the subject of methodological experiments. Although those experiments have rarely involved targeting, they can be used to simulate the effect of targeting, by comparing effects between sample subgroups. (The studies of Jäckle (2015) and Pratt et al. (2015) both explicitly state that a randomised rather than targeted assignment to treatment was used in order to be able to identify how best to target in future.) Thus, useful evidence about the likely effects of targeting could be gleaned from secondary analysis of existing experimental data, without the need to conduct new experiments. Such analysis, to inform future targeted designs, would surely be a worthwhile task for the survey research community. It may even be possible to infer relevant effects from natural 'experiment' experiments where procedures were changed over time or varied between sample subgroups for other reasons (for example, differences in regional practices), so long as the underlying survey conditions are reasonably constant.

Improving the relationship between survey costs and survey errors through targeting can be achieved in many ways. It can involve redistributing a fixed budget in a way that reduces one or more source of error, or that reduces one source of error to an extent that outweighs a concomitant increase in another source of error. Alternatively, it could involve slightly increasing the costs to bring about a substantial reduction in error, or indeed reducing the costs without affecting error. It need not involve affecting both the costs and the errors: affecting just one of the two can be enough to change the relationship between them. Though some targeting methods certainly increase costs – for example an additional intervention for a particular subgroup, such as a between-wave telephone call – this need not be the case. Targeting can be used as a tool to better allocate scarce survey resources in order to improve survey outcomes, as demonstrated in Luiten and Schouten (2013). However, it should be noted that even cost-neutral targeted designs may come at the price of increased risk, compared to a standardised design. The risk is that of incorrect administration of the targeted feature, such that some sample members receive a variant that was not the intended one. The consequences may or may not be innocuous, depending on the nature of the treatment.

A distinction can also be made between approaches to targeting that involve introducing a new feature or procedure that would not otherwise have been deployed on the survey and those that only involve modifying an existing feature. The latter are generally cheaper and easier to implement and may be worth considering even if the evidence of a positive effect is weak (provided there is little risk of a negative effect). Producing variants of written communications with sample members, as in Fumagalli et al. (2013), Cleary and Balmer (2015) and Lynn (2016), is a good example. If it is already planned to send an email, letter or leaflet to each sample member, targeting the wording or design of that communication will not affect the cost of administration. The only costs will be those of design and production. Interventions that rely on interviewers changing their calling behaviours in suggested ways, on the other hand, may not achieve the predicted outcomes unless ways can be found to improve on the low levels of interviewer compliance reported by Wagner (2013) and Kreuter and Müller (2015).

Table 1
Studies of targeted design features, classified by agent of change, change mechanism, and affected outcome

Agent of change	Change mechanism	Affected outcome	Targeted design feature	Studies
Respondent	Motivation	Co-operation	Respondent communication	Fumagalli, Laurie, and Lynn (2013); Cleary and Balmer (2015); Lynn (2016)
			Respondent incentives	Kay, Boggess, Selvavel, and McMahon (2001); Bureau of Labor Statistics (undated); Zukerberg, Hall, and Henly (2007); Jessop and Oksala (2014)
		Location	Extra contacts	Lynn (2012); Fumagalli, Laurie, and Lynn (2013); Cleary and Balmer (2015)
	Burden	Co-operation	Data collection modes	Luiten and Schouten (2013); Rosen et al. (2014); Al Baghal and Lynn (in progress)
Interviewer	Motivation	Contact & co-operation	Interviewer incentives	Peytchev, Riley, Rosen, Murphy, and Lindblad (2010); Calderwood, Carpenter, and Cleary (2013)
	Barriers	Location	Field priority	Walejko (2015)
		Contact	Field priority	Calderwood, Cleary, Flore, and Wiggins (2012)
			Call scheduling algorithm	Luiten and Schouten (2013); Kreuter and Müller (2015)
		Co-operation	Interviewer allocation	Luiten and Schouten (2013)
			Persuasion scripts	Lipps (2012)

6 Looking Forward

There is currently much interest in targeted designs. Apart from studies of respondent incentives, all the studies summarised in section 4 above and listed in table 1 have been published since 2010. Researchers have been inventive in the range of design features that have been applied in a targeted fashion. Furthermore, several experimental studies have demonstrated that desired outcomes can be achieved, both in terms of improving response rates and in terms of improving sample balance. However, it remains the case that few surveys implement targeting routinely and that few design features are ever targeted as a means of tackling nonresponse. Targeting of respondent incentives is the only example that is widespread. However, there are reasons to suppose that this may change.

A significant barrier to rapid adoption of targeted survey design features is that survey organisations are not currently set up to implement such features without disruption to their usual procedures. For many design features a modest investment would be needed to modify organisation-wide systems to allow the possibility of multiple targeted variants. Once the systems are modified in this way, it should be straightforward to toggle the option on or off for any particular survey, and to set the parameters of the variants, enabling incorporation of targeted design features without affecting the survey timetable.

It may soon be the case that rather than asking themselves whether they should use a targeted design, survey researchers will instead be asking which features of their survey they should design in a targeted way. Targeted design features may become much more commonplace, particularly in longitudinal surveys, where the conditions for successful targeting can usually be met. Such designs may help researchers to achieve quality objectives even in the face of tightened budgets, by achieving a better balance between survey costs and survey errors.

There are at least two directions in which one might anticipate that research into targeted design features might develop

in future. One is in the direction of increased sophistication of the targeted treatments. For example, targeting to increase contact rates amongst the hardest-to-contact subgroups could involve call scheduling algorithms that take into account predicted probabilities of contact in various time slots in a probabilistic way, rather than relying on a crude dichotomy into cases that deserve priority attention and those that do not. Similarly, attempts to maintain updated contact details for sample members at greatest risk of moving could involve a range of different tracking and communication activities for subgroups at risk for different reasons, such as young adults still living with their parents, young unmarried professionals in private rented accommodation, persons approaching retirement age, and so on.

The second direction in which research into targeted design features might develop in future is towards increased sophistication of the objectives. To date, the objective has generally been to improve one or, occasionally, more of the location rate, contact rate or co-operation either simply to improve the response rate or to improve sample balance (as a proxy for nonresponse bias). Instead, it should be possible to consider the location propensity, contact propensity and co-operation propensity of each sample member simultaneously in order to identify an overall design package to maximise participation propensity. This would recognise, for example, that the value of locating a hard-to-locate sample member is not independent of the conditional propensity to participate once located. Future strategies may even go beyond nonresponse error and adopt a total survey error approach (Lynn & Lugtig, 2017) to minimise the overall combined error from multiple sources, for example nonresponse error and measurement error.

Acknowledgements

The author is part-funded by the UK Economic and Social Research Council (award number ES/H029745/1 for "Understanding Society: the UK Household Longitudinal Study"). This work was first presented at a seminar in Copenhagen, Denmark, on "Present challenges facing survey researchers", organized by the Danish Society for Survey Research, in February 2016. Versions were also presented at the International Workshop on Panel Survey Methods in Berlin, in June 2016, and the annual conference of the (UK) Social Research Association in London, in December 2016.

References

Al Baghal, T. & Lynn, P. (in progress). Targeted allocation to mode based on multiple objectives. In progress.

Al Baghal, T. & Lynn, P. (2015). Using motivational statements in web-instrument design to reduce item-missing rates in a mixed-mode context. *Public Opinion Quarterly*, *79*(2), 568–579.

Beaule, A., Brown, C., Campbell, F., Dascola, M., Freedman, V., Insolera, N., ... Warra, J. (2015). *PSID Main Interview User Manual: Release 2015*. Ann Arbor: University of Michigan. Retrieved from http://psidonline.isr.umich.edu

Beaumont, J.-F., Bocci, C., & Haziza, D. (2014). An adaptive data collection procedure for call prioritization. *Journal of Official Statistics*, *30*(4), 607–621.

Bennett, D. J. & Steel, D. (2000). An evaluation of a large-scale cati household survey using random digit dialling. *Australian and New Zealand Journal of Statistics*, *42*(3), 255–270.

Bethlehem, J., Cobben, F., & Schouten, B. (2011). *Handbook of nonresponse in household surveys*. Wiley.

Bureau of Labor Statistics. (undated). *National Longitudinal Survey of Youth 1997: Index to the NLSY97 Cohort*. Retrieved from https://www.nlsinfo.org/content/cohorts/nlsy97

Calderwood, L., Carpenter, H., & Cleary, A. (2013, September). *Experiments to improve response rates among likely refusers on longitudinal surveys: Does assigning better interviewers and paying interviewer incentives work?* Paper presented at the International Workshop on Household Survey Nonresponse, London UK.

Calderwood, L., Cleary, A., Flore, G., & Wiggins, R. T. (2012, July). *Using response propensity models to inform fieldwork practice on the fifth wave of the Millenium Cohort Study*. Paper presented at the International Panel Survey Methods Workshop, Melbourne, Australia.

Campanelli, P., Sturgis, P., & Purdon, S. (1997). *Can you hear me knocking? and investigation into the impact of interviewers on survey response rates*. The Survey Methods Centre at SCPR, London, GB. Retrieved from http://eprints.soton.ac.uk/80198/

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, *12*, 389.

Cleary, A. & Balmer, N. (2015). Fit for purpose? The impact of between-wave engagement strategies on response to a longitudinal survey. *International Journal of Market Research*, *57*(4), 533–554.

Couper, M. P. & Ofstedal, M. B. (2009). Keeping in Contact with Mobile Sample Members. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 183–203). Chichester UK: Wiley.

Eggs, J. & Jäckle, A. (2015). Dependent interviewing and sub-optimal responding. *Survey Research Methods*, *9*(1), 15–29.

ESS Sampling Expert Panel. (2014). *Sampling for the European Social Survey Round VII: Principles and Requirements, 2nd version*. Retrieved from http://www.

europeansocialsurvey . org / docs / round7 / methods / ESS7%5C_sampling%5C_guidelines.pdf

Fumagalli, L., Laurie, H., & Lynn, P. (2013). Experiments with methods to reduce attrition in longitudinal surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(2), 499–519.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, *56*(4), 475–495.

Groves, R. M. & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(3), 439–457.

Haan, M. & Ongena, Y. (2014). Tailored and targeted designs for hard-to-survey populations. In R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hard-to-Survey Populations* (pp. 555–574). Cambridge UK: Cambridge University Press.

Jäckle, A. (2009). Dependent interviewing: A framework and application to current research. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 93–111). Chichester UK: Wiley.

Jäckle, A. (2015). Going online with a face-to-face household panel: effects of a mixed mode design on item and unit nonresponse. *Survey Research Methods*, *9*(1), 57–70.

Jessop, C. & Oksala, A. (2014). *UK Household Longitudinal Study: Wave 4 Technical Report*. London: NatCen Social Research. Retrieved from https : // www . understandingsociety.ac.uk/documentation/mainstage/technical-reports

Johansson, A., Lundquist, P., Westling, S., & Durrant, G. B. (2015, November). *Modelling long call sequences and final outcome in the swedish labour force survey to reduce the number of unproductive calls*. Paper presented at the Adaptive Survey Design Workshop, Manchester, UK. Retrieved from http : // www . cmist . manchester . ac . uk / research / projects / baden / people - and-events/conference-2015/

Kaplowitz, M. D., Lupi, F., Couper, M. P., & Thorp, L. (2012). The effect of invitation design on web survey response rates. *Social Science Computer Review*, *30*(3), 339–349.

Kay, W. R., Boggess, S., Selvavel, K., & McMahon, M. F. (2001). The use of targeted incentives to reluctant respondents on response rate and data quality. *Proceedings of the Annual Meeting of the American Statistical Association*.

Kreuter, F. & Müller, G. (2015). A note on improving process efficiency in panel surveys with paradata. *Field Methods*, *27*(1), 55–65.

Laurie, H. & Lynn, P. (2009). The use of respondent incentives on longitudinal surveys. In P. Lynn (Ed.), *Methodology of Longitudinal Surveys* (pp. 205–233). Chichester UK: Wiley.

Lipps, O. (2012). Using information from telephone panel surveys to predict reasons for refusal. *Methoden-Daten-Analysen*, *6*(1), 3–20.

Londra, M. & Calderwood, L. (2006). *Millenium Cohort Study Second Survey: CAPI Questionnaire Documentation Version 1*. London: Centre for Longitudinal Studies. Retrieved from www.cls.ioe.ac.uk

Luiten, A. & Schouten, B. (2013). Tailored fieldwork design to increase representative household survey response: an experiment in the survey of consumer satisfaction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 169–189.

Lynn, P. (2012, September). *Failing to locate panel sample members: minimising the risk*. Paper presented at the International Workshop on Household Survey Nonresponse, Ottawa, Canada.

Lynn, P. (2014). Targeted response inducement strategies on longitudinal surveys. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds.), *Improving Survey Methods: Lessons from Recent Research* (pp. 322–338). Abingdon UK: Psychology Press.

Lynn, P. (2016). Targeted appeals for participation in letters to panel survey members. *Public Opinion Quarterly*, *80*(3), 771–782.

Lynn, P. & Lugtig, P. (2017). Total survey error for longitudinal surveys. In P. Biemer, E. De Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, . . . B. West (Eds.), *Total Survey Error in Practice* (Chap. 14). Hoboken NJ: Wiley.

Lynn, P. & Sala, E. (2006). Measuring change in employment characteristics: the effects of dependent interviewing. *International Journal of Public Opinion Research*, *18*(4), 500–509.

Mathiowetz, N. A. & McGonagle, K. A. (2000). An assessment of the current state of dependent interviewing in household surveys. *Journal of Official Statistics*, *16*(4), 401–418.

McGonagle, K. A., Couper, M. P., & Schoeni, R. F. (2011). Keeping track of panel members: an experimental test of a between-wave contact strategy. *Journal of Official Statistics*, *27*(2), 319–338.

Moore, J. C. (2007, December). *Seam bias in the 2004 SIPP Panel: much improved, but much bias still remains*. Paper presented at the US Census Bureau / PSID Event History Calendar Research Conference. Retrieved from http : / / psidonline . isr . umich . edu / Publications / Workshops / ehc - 07papers / Seam % 20Bias % 20in % 20the % 202004 % 20SIPP % 20Panel . pdf.

Morton-Williams, J. (1993). *Interviewer Approaches*. Aldershot: Dartmouth.

Oudejans, M. & Scherpenzeel, A. (2012). *Especially for you: motivating respondents in an internet panel by offering tailored questions*. Paper Presented at the 8th International Conference on Social Science Methodology (RC33), Sydney.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010). Reduction of nonresponse bias in surveys through case prioritization. *Survey Research Methods*, *4*(1), 21–29.

Pratt, D., Cominole, M., Copello, E., Peytchev, A., Peytcheva, E., Rosen, J., & Wilson, D. (2015, November). *Examination of interventions during data collection to increase response and sample representativeness: a field test experiment and simulation*. Paper presented at the Adaptive Survey Design Workshop, Manchester, UK. Retrieved from http://www.cmist.manchester.ac.uk/research/projects/baden/people-and-events/conference-2015/

Rosen, J. A., Murphy, J., Peytchev, A., Holder, T., Dever, J. A., Herget, D. R., & Pratt, D. J. (2014). Prioritizing low-propensity sample members in a survey: implications for nonresponse bias. *Survey Practice*, *7*(1).

Sala, E., Knies, G., & Burton, J. (2014). Propensity to consent to data linkage: experimental evidence on the role of three survey design features in a UK longitudinal panel. *International Journal of Social Research Methodology*, *17*(5), 455–473.

Schober, M. F. & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, *61*(4), 576–602.

Schouten, B., Cobben, F., Lundquist, P., & Wagner, J. (2016). Does more balanced survey response imply less nonresponse bias? *Journal of the Royal Statistical Society Series A (Statistics in Society)*, *179*(3), 727–748.

Singer, E. (2002). The use of incentives to reduce nonresponse in household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 163–177). New York: Wiley.

Suchman, L. & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association*, *85*(409), 232–53.

VanGeest, J. B., Johnson, T. P., & Welch, V. L. (2007). Methodologies for improving response rates in surveys of physicians. *Evaluation & the Health Professions*, *30*(4), 303–321.

Verma, V. & Betti, G. (2010). Data accuracy in EU-SILC. In A. B. Atkinson & E. Marlier (Eds.), *Income and Living Conditions in Europe* (pp. 57–77). Luxembourg: Publications Office of the European Union.

Wagner, J. (2008). *Adaptive survey design to reduce nonresponse bias*. PhD thesis, University of Michigan, USA. Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/60831

Wagner, J. (2013). Adaptive contact strategies in telephone and face-to-face surveys. *Survey Research Methods*, *7*(1), 45–55.

Walejko, G. (2015, November). *Prioritizing cases to increase sample representativeness in the survey of income and program participation*. Paper presented at the Adaptive Survey Design Workshop, Manchester, UK. Retrieved from http://www.cmist.manchester.ac.uk/research/projects/baden/people-and-events/conference-2015/

Watson, N. & Wilkins, R. (2012). *The impact of computer-assisted interviewing on interview length*. HILDA Project Discussion Paper 1/12. Retrieved from http://www.melbourneinstitute.com/hilda/biblio/HILDA_discussion_papers.html

Wolff, P., Montaigne, F., & Rojas González, G. (2010). Investing in statistics: EU-SILC. In A. B. Atkinson & E. Marlier (Eds.), *Income and Living Conditions in Europe* (pp. 37–55). Luxembourg: Publications Office of the European Union.

Zagorsky, J. L. & Rhoton, P. (2008). The effects of promised monetary incentives on attrition in a long-term panel survey. *Public Opinion Quarterly*, *72*(3), 502–513.

Zukerberg, A., Hall, D., & Henly, M. (2007). *Money can buy me love: experiments to increase response through the use of monetary incentives.* Washington DC: U.S. Census Bureau.