Collection of Delinquent Fines: An Adaptive Randomized Trial to Assess the Effectiveness of Alternative Text Messages

Laura C. Haynes
Donald P. Green
Rory Gallagher
Peter John
David J. Torgerson

Abstract

The collection of delinquent fines is a vast and ongoing public administration challenge. In the United Kingdom, unpaid fines amount to more than 500 million pounds. Managing noncompliant accounts and dispatching bailiffs to collect fines in person is costly. This paper reports the results of a large randomized controlled trial, led by the UK Cabinet Office's Behavioural Insights Team, which was designed to test the effectiveness of mobile phone text messaging as an alternative method of inducing people to pay their outstanding fines. An adaptive trial design was used, first to test the effectiveness of text messaging against no treatment and then to test the relative effectiveness of alternative messages. Text messages, which are relatively inexpensive, are found to significantly increase average payment of delinquent fines. We found text messages to be especially effective when they address the recipient by name. © 2013 by the Association for Public Policy Analysis and Management.

INTRODUCTION

The collection of delinquent fines is a vast and ongoing public administration challenge. In the United Kingdom, the Ministry of Justice handles over 2 million criminal cases, 1.8 million civil claims, 150,000 family law disputes, and 800,000 tribunal cases annually. Each year, the Ministry imposes over 1 million new court fines, with a value of over £350 million. However, only 50 percent of these are collected within six months (National Audit Office, 2007), and in 2011, the value of outstanding court fines was estimated to be over £600 million (United Kingdom Ministry of Justice, 2011; United Kingdom Ministry of Justice, Public Account Committee, 2012). Recovery of outstanding court fines requires resources. Staff time is required to follow up with debtors by phone; failure to secure funds by phone causes a case to be referred to a bailiff, who must visit a debtor's home and in some cases seize property. In 2012, there were almost 2,000 certified bailiffs registered in England and Wales, and the Ministry of Justice estimates they are responsible for enforcing approximately 580,000 criminal fines annually. Debtors referred to the bailiffs incur a minimum administration fee of £75,[1] although this fee rises if additional visits to the residence are required. While cases referred to the bailiffs do not directly impose

---

[1] https://consult.justice.gov.uk/digital...bailiff.../trasnforming-bailiff-ia.pdf.

costs on the state, the indirect costs can be significant, as a substantial proportion of such cases are not settled and return to Her Majesty's Courts and Tribunals Service (HMCTS)—the agency that administers the collection of fines in the UK—for further processing.

In an effort to develop cost-effective strategies for improving collection rates, the Behavioural Insights Team (UK Cabinet Office) and HMCTS, initiated a series of randomized trials designed to test the effectiveness of low-cost fine collection strategies. Although several previous experiments have assessed government efforts to induce tax compliance through audits (Kleven et al., 2010), warning letters threatening audits (Iyer, Reckers, & Sanders, 2010; Slemrod, Blumenthal, & Christian, 2001), and letters pleading with citizens to pay their fair share (Fellner, Sausgruber, & Traxler, 2013), to our knowledge, this is the first time that field experiments have been used to evaluate the effectiveness of strategies for collecting delinquent fines.

Our study also breaks new ground insofar as it assesses the effects of text messaging through mobile phones as a means of collecting fines. Brief text messages were judged to have promise after initial fieldwork indicated that debtors often fail to open warning letters or to understand their instructions. Moreover, text messages represent a low-cost means of prompting payment from large numbers of people with outstanding fines. In the months leading up to our study, HMCTS used this method on an ad hoc basis, deploying what we later refer to as the "standard" treatment. Because text messages are delivered by automated systems, the content of these messages may be easily and inexpensively varied, even to the point where messages are customized to each recipient. The central research questions behind this study are whether text messages induce recipients to pay their outstanding fines in a timely manner and, if so, which messages are most effective.

Prior research conducted in a range of different contexts strongly suggests that text messaging has the potential to influence behavior. Text messages have been shown to increase personal savings (Karlan et al., 2010), rates of voter turnout (Dale & Strauss, 2009), energy conservation (Gleerup et al., 2010), smoking cessation (Free et al., 2011), and positive health behaviors more generally (Fjeldsoe, Marshall, & Miller, 2009). The challenge is to formulate messages that will extract payments from those who have defied a court order to pay.

Drawing on the literatures on tax compliance and health psychology, we devised a series of experimental messages that incorporated three ingredients. The first is a strong signal to the recipient that noncompliance has been noticed and that there is now a high risk of punitive action. People who fail to respond to warning letters are arguably less sensitive to threats of this kind than the general population of taxpayers studied by Slemrod, Blumenthal, and Christian (2001), Kleven et al. (2010), and Iyer, Reckers, and Sanders (2010). However, the Fellner, Sausgruber, and Traxler (2013) study, which looked at Austrians who failed to pay their television tax, found that threats of punishment were effective in inducing compliance. Second, timing is likely to play an important role here. The text demanded immediate action (and provided a phone number to call) and warned that failure to act would cause the case to be referred to a bailiff. A third ingredient is customization. Sending a message that tells each recipient the amount of his or her outstanding fine signals the government's capacity to retrieve and act upon information that could lead to punishment. Moreover, by comparison to a generic message, a message customized in this way may be more likely to attract the recipient's attention (Dijkstra, 2005). Another form of customization is to address each recipient by name. A wealth of psychological evidence suggests the special power of names in attracting attention (Bargh, 1982). The so-called "cocktail party effect" (Cherry, 1953), whereby people filter out competing stimuli and refocus their attention when their name is mentioned, has been shown to operate even when names appear in printed text (Shapiro, Caldwell, &

Sorensen, 1997). In sum, customized treatments were intended to emphasize the likelihood of punitive action and attract the recipient's attention.

This paper is organized as follows. We begin by describing an adaptive randomized trial conducted in the southeast of England, where we tested the impact of different text messages on the amounts subsequently paid by those who had previously failed to pay their court-ordered fines. Next we describe our statistical framework for analyzing the results given the fact that some of the text messages that HMCTS sent were found to be undeliverable. Finally, we present the results from each phase of the trial, which suggest that text messaging is highly effective in generating immediate payment of fines, especially when the messages address the recipient by name.

EXPERIMENTAL DESIGN

Trial Design

The trial design was a multiarm, multistage, adaptive-randomized trial. In other words, the trial began with several treatment groups and sought to winnow out ineffective treatments. Adaptive designs are sometimes used in medicine to reduce the number of participants who are exposed to an inferior treatment. They may also be used to increase the efficiency of trials. For example, in the STAMPEDE trial (Sydes et al., 2012), men with prostate cancer were initially randomized to several treatment groups. Treatments that did not appear to be producing any benefit were closed, and more participants were subsequently randomized to remaining treatment groups. Our trial design adopted a similar approach. At the outset, it was suspected that text messaging would be superior to no messaging. Consequently, once sufficient numbers of participants had been randomized to the no-text control condition to support this hypothesis with sufficient confidence, this treatment group was eliminated, and the remaining participants were allocated to the four remaining treatment groups. This design allows a robust estimate of a treatment effect against an untreated control but does not waste participants by continuing to randomize them to the no-treatment condition.

Sample Allocation

The trial was conducted in three regions in the southeast of England. The population consisted of individuals for whom HMTCS held a mobile phone number and whose failure to pay a court-ordered fine had led to an escalation of their case to "distress warrant" status. A distress warrant is a court order that empowers a bailiff to recover the debt directly, often through the confiscation and sale of possessions. The reasons for receiving a distress warrant varied. Some subjects failed to pay according to the timetable set up by the HMCTS; other distress warrants were issued after unsuccessful arrest warrants were returned by court officers and a new address was found. All debtors in the sample had received a minimum of one written warning of the consequences of nonpayment of the outstanding fine, which included arrest and confiscation of their property.

At the beginning of each week, a list was compiled of cases that had reached distress warrant status during the preceding seven days. Those cases for which a mobile number was not held were removed from the sample, and remaining cases were allocated randomly to five experimental conditions during the Phase 1 trial (January 2012 through early February 2012) or four experimental conditions during the Phase 2 trial (February 2012 through April 2012). During both phases, allocation was conducted by a HMCTS analyst using simple random assignment with equal probabilities of assignment to each experimental group. In Phase 1, a total of 1,817

Table 1. Messages associated with each treatment condition.

Text condition (abbreviation)	Text message
None (NONE)	[No text message sent.]
Standard (STANDARD)	You have not paid your fine. Pay immediately or a warrant will be issued to the bailiffs. Call 03007909901 quote ref [number] div [number].
Personalized name (PERSONAL)	[Name], you have not paid your fine. Pay immediately or a warrant will be issued to the bailiffs. Call 03007909901 quote ref [number] div [number].
Personalized amount (AMOUNT)	You have not paid your fine of £[amount]. Pay immediately or a warrant will be issued to the bailiffs. Call 03007909901 quote ref [number] div [number].
Personalized name and amount (PERSONAL/AMOUNT)	[Name], you have not paid your fine of £[amount]. Pay immediately or a warrant will be issued to the bailiffs. Call 03007909901 quote ref [number] div [number].

subjects were randomly allocated; another 3,633 subjects were randomly allocated in Phase 2. Multinomial logistic regression confirms that the random assignments in each phase were not significantly predicted by the subjects' age, number of previous distress warrants, or gender. In Phase 1, the likelihood-ratio chi-square test had a *p* value of 0.46; in Phase 2, the *p* value was 0.52.

Experimental Treatments

On the Monday after their cases escalated to distress warrant status, HMCTS sent individuals in the text message treatment groups a text (short message service [SMS]) to their mobile phones. These messages alert the owner that the message has arrived; the owner presses a button to see the text on the screen of the phone. Recipients of the text message viewed the sender as "HMCTS." In the event that the message was undeliverable, the sender received a notification, which we used to classify text messages as "delivered" or "not delivered."

As shown in Table 1, each message conveyed the same core information. Recipients were reminded about their unpaid fines, warned that failure to pay would result in a warrant, instructed to call a payment hotline number, and given the reference identification number. The experimental variations on the standard treatment (see Table 1) were the personalized (PERSONAL) condition, in which the message was preceded by the recipient's name; the personalized amount (AMOUNT) condition, in which the recipient was reminded of the total value of the outstanding fine; and the personalized name and amount (PERSONAL/AMOUNT) condition, which included both of these elements.

Outcomes

Using HMCTS records, outcomes for each subject were measured over a period of seven days in both the Phase 1 and 2 trials.[2] Recipients who failed to pay within

---

[2] The same staff analyst who conducted the random assignment also assembled the outcome data. A computer script was used to automate assignment, but IT systems required manual checking of accounts to record any payment activity in the seven days since treatment.

seven days of receiving the text had their cases referred to the bailiffs. In the interest of focusing attention on the most policy-relevant outcome, the analysis presented below considers only the actual amount that the individual paid (ignoring promises to pay that were not accompanied by actual payment). This outcome was observed for all subjects in both phases of the trial.

IDENTIFICATION AND ESTIMATION OF CAUSAL PARAMETERS OF INTEREST

The treatment effect of a given intervention (e.g. the PERSONAL text treatment) may be defined as the difference between two potential outcomes (Rubin, 2005). One potential outcome, denoted $Y_i(0)$, is the amount that subject i would pay in fines if no treatment were administered; another potential outcome, denoted $Y_i(1)$, is the amount that subject i would pay in fines if he or she receives the PERSONAL text. The average treatment effect, or ATE, is the average of $Y_i(1) - Y_i(0)$ across the entire subject pool. Because we observe each subject in either a treated or untreated state, it is impossible to observe a subject's treatment effect, $Y_i(1) - Y_i(0)$. Random assignment, however, allows us to use the average outcome in the treatment group to estimate the average $Y_i(1)$ and the average outcome in the control group to estimate the average $Y_i(0)$; these two pieces of information allow for unbiased estimation of the ATE (Gerber & Green, 2012, chapter 2).

One complication that arises in the context of an experiment that uses texting as a treatment is that only some of the subjects who are sent a text actually receive it. To formalize our description of the statistical problem, let z_i be 1 when subject i is assigned to the treatment group and let z_i be 0 when subject i is assigned to the control group. In order to distinguish assigned from actual treatment, we let d_i be 1 when subject i is actually treated and let d_i be 0 when subject i is not treated. In our study, 54.5 percent of the 5,084 texts that were sent during the two experimental phases were actually delivered. In other words, we observe 2,770 subjects for whom $z_i = 1$ and $d_i = 1$ and another 2,314 subjects for whom $z_i = 1$ and $d_i = 0$. No one in our study received the treatment when assigned to the control group ($N = 366$). Following the terminology of Angrist, Imbens, and Rubin (1996), our subject pool may be said to consist of two latent groups: never-takers (those who never receive the treatment regardless of their experimental assignment because they cannot receive text messages) and compliers (those who would receive a text message if assigned to the treatment group).

When only some of the subjects assigned to the treatment group actually receive treatment, a randomized experiment cannot recover the average treatment effect defined above (Angrist, Imbens, & Rubin, 1996). Instead, an experiment that fails to treat some portion of the assigned treatment group may be used to estimate two alternative estimands, the average intention-to-treat (ITT) effect and the complier average causal effect (CACE). The average ITT effect is the average effect of assigning subjects to the treatment group, regardless of whether they are actually treated. This estimand may be useful from a policy standpoint because it gives a sense of the net effect of a program given limitations of implementation. In the context of the present study, the ITT represents the effect of attempting to send text messages to those who have reached DW status. The CACE is the average treatment effect among a subset of the subject pool, compliers. Unfortunately, we cannot learn anything about treatment effects among never-takers (those who cannot receive texts) because we never observe them in their treated state.

Estimation of the ITT and CACE is straightforward. Unbiased estimates of the ITT are easily obtained by subtracting average outcomes in the assigned control group from average outcomes in the assigned treatment group. Under the assumption that assignment on its own has no effect on outcomes, consistent estimates of the CACE

are obtained by dividing the estimated ITT by the fraction of subjects in the assigned treatment group who actually receive the treatment (Gerber & Green 2012, chapter 5). In order to see the intuition behind this estimator, notice that expected outcomes in the assigned control group may be expressed as a weighted average of expected untreated outcomes among compliers and never-takers, where the weights[3] are the proportions of compliers and never-takers:

$$E[Yi \backslash Zi = 0] = E[Yi(0)/compliers] \cdot Pr[compliers] + E[Yi(0)never\text{-}takers]$$

$$\cdot \Pr[never\text{-}takers]. \tag{1}$$

Since treatment assignment is random, the proportions of compliers and never-takers are, in expectation, identical in both treatment and control groups. Therefore, expected outcomes in the treatment group may be expressed as a weighted average using the same weights:

$$E[Y_i \backslash Z_i = 1] = E[Y_i(1)/compliers] \cdot Pr[compliers] + E[Y_i(0)never\text{-}takers]$$

$$\cdot \Pr[never\text{-}takers]. \tag{2}$$

Subtracting equation (1) from equation (2) and dividing by the proportion of compliers gives

$$E[Y_i(1)|compliers] — E[Y_i(0)/compliers] = E[(Y_i(1) — Y_i(0))/compliers], \tag{3}$$

which is the CACE. In practice, the estimator for equation (3) uses the observed average outcome in the control group to estimate equation (1), the observed average outcome in the treatment group to estimate equation (2), and the observed rate of treatment in the assigned treatment group to estimate Pr[*Complier*]. This estimator is equivalent to instrumental variables regression in which outcomes are regressed on actual treatment, which is instrumented by assigned treatment (Angrist, Imbens, & Rubin, 1996).

Another complication associated with our experimental design is the presence of multiple treatment groups. Fortunately, this complication poses no special estimation problems. Suppose we seek to estimate the CACE of receiving the PERSONAL treatment as opposed to the STANDARD treatment. Because both groups are sent text messages at the same time and under identical conditions, an equivalent set of compliers receives each type of text message (Gerber et al., 2010). Estimating the CACE is simply a matter of comparing average outcomes among those who receive the PERSONAL treatment to the average outcomes among those who receive the STANDARD treatment. In contrast to the instrumental variables estimator described above, which is consistent but biased, this simple difference-in-means estimator is unbiased.

A final complication associated with our outcome data is that we observe a large group of subjects who pay nothing in fines and a relatively small group who pay amounts ranging from 2 pounds to 615 pounds. For example, no fine was collected from 73.2 percent of the 2,770 subjects who actually received a treatment text. This skewed distribution presents two types of estimation concerns. First, it complicates the use of linear regression when covariates are included as right-hand-side predictors; the inclusion of covariates may lead to negative (and therefore inadmissible) predicted values. Rather than introduce other modeling assumptions (such as

---

[3] The weights add up to 1.0 because the probability of being a complier is the complement of the probability of being a never-taker.

normally distributed disturbances, as in the Tobit model), we take a nonparametric approach and simply compare average outcomes across different experimental groups. A second complication is heteroskedasticity, since a treatment that increases the average payment also tends to increase the variance in payments. In order to test hypotheses in a manner that is robust to heteroskedasticity and skewness, we use randomization inference to obtain exact *p* values when testing the sharp null hypothesis of no effect for any subject (Gerber & Green, 2012, chapter 3). This procedure boils down to simulating the sampling distribution from 100,000 hypothetical random assignments under the null hypothesis that $Y_i(1) = Y_i(0)$ for all subjects. We also use randomization inference when comparing different messages. In this case, we test the null hypothesis that $Y_i(A) = Y_i(B)$ for those who actually received messages A or B.

RESULTS

Our experimental design consisted of two phases: In Phase 1, we tested a series of alternative text treatments against a control condition in which no text was sent. The null hypothesis is that text messages fail to increase the payment of delinquent fines, which implies the use of one-tailed tests. In terms of point estimation, this phase of the experiment allowed us to gauge the overall effectiveness of text messaging and to assess tentatively the relative effectiveness of different types of messages.[4] In Phase 2, the NO TEXT group was eliminated from the design, and the aim was to sharpen our estimates of the relative effectiveness of alternative messages. Two-sided tests are used to test the null hypothesis that subjects respond to alternative messages in the same way.

Phase 1 Trial

Table 2 reports the average amount paid by subjects in each of the assigned experimental groups. These figures, which make no allowance for whether texts were actually received, permit us to estimate the ITT of each of the text treatments vis-à-vis the control condition in which no text was sent. In the NO TEXT condition, the average payment was £4.46. By contrast, average payment in the PERSONAL condition was £12.87, an £8.41 or 189 percent increase (one-tailed $p < 0.001$). The AMOUNT condition generated a £6.07 increase over the NO TEXT baseline (one-tailed $p = 0.007$). When the text message combined the elements of PERSONAL and AMOUNT, the increase was £7.28 (one-tailed $p = 0.003$), which was smaller than PERSONAL alone but larger than AMOUNT alone. The weakest performer was the STANDARD text message, which included neither the amount nor the personal information. Here, the increase over the NO TEXT condition was £4.16 (one-tailed $p = 0.034$).[5]

Because we know the proportion of subjects who actually received the intended text messages, we can estimate the complier average causal effect for each treatment condition. Overall, 55.3 percent of the Phase 1 subjects in the four text treatment conditions actually received their text. Rates of successful message delivery across the four treatment conditions vary slightly but no more than would be expected

---

[4] The Phase 1 data were analyzed on February 2, 2012, after 200 observations were gathered in each group, at which point the results showed a significant effect of personalization (two-tailed = 0.003) relative to the NO TEXT condition. At that point it was decided to end Phase 1 during the last week of February.
[5] These results change only trivially when covariate adjustment is used to control for the date on which the text message was sent or the recipient's age, gender, or prior number of distress warrants.

Table 2. Results of the Phase 1 trial, by experimental condition.

Phase 1	N	Avg. amount paid (in pounds)	95% Confidence interval of average amount paid[a]	Percentage of subjects reached by text	Estimated ITT effect	95% Confidence interval of ITT effect[b]	Estimated CACE
NONE	366	4.46	[1.97, 7.47]	0.0			
STANDARD	361	8.62	[5.48, 12.61]	56.0	4.16*	[-0.30, 8.60]	7.43
AMOUNT	364	10.53	[6.76, 14.83]	52.5	6.07**	[1.22, 10.91]	11.57
PERSONAL	362	12.87	[9.00, 17.21]	55.5	8.41***	[3.44, 13.38]	15.14
PERSONAL/ AMOUNT	364	11.74	[7.52, 16.99]	57.4	7.28****	[1.91, 12.69]	12.68

The ITT (intention-to-treat) and CACE (complier-average causal effect) are calculated in comparison to the NONE condition.

*One-tailed value is 0.034.

**One-tailed value is 0.007.

***One-tailed value 0.001.

****One-tailed value is 0.003.

[a]Calculated using 100,000 bootstrap samples.

[b]Calculated using the package in R. See Aronow and Samii (2012) and Gerber and Green (2012), Chapter 3, for more on this procedure.

by random sampling variability (a chi-square test of equal proportions across the four conditions has a *p* value of 0.59). Regression analysis predicting successful delivery of the text message reveals significant relationships for gender, age, and prior history of DWs. Males, young people, and (perhaps counter intuitively) those with no prior DW history were contacted at lower rates. The latter finding may reflect the fact that those with prior DWs may have been more likely to supply a working mobile number in the course of repeated interactions with HMCTS. Although these three relationships are each highly significant ($p < .001$), they jointly predict only 4 percent of the variance in compliance.

Dividing the estimated ITT effect for each experimental condition by the proportion who received the text message in each condition provides a consistent estimator of each treatment's average effect among compliers. These figures are presented in the rightmost columns of Table 2. The estimates indicate that for compliers, the PERSONAL treatment raises the average contribution by £15.15. The next largest effect (£12.69) is associated with the PERSONAL/AMOUNT combination. AMOUNT alone generates an average contribution of £11.57 more than NO TEXT. The weakest estimated CACE (£7.43) is observed among those who received the STANDARD text.

The results of the Phase 1 study clearly indicate that text messaging is an effective intervention. Putting aside variations in message content, we find that sending a text boosted the average amount paid in fines by 145 percent (from £4.46 to £10.94). When we restrict attention to the 803 subjects who were reachable by SMS, the average amount paid in fines rises by 210 percent.[6] The results of the Phase 1 trial leave little doubt as to whether texting induces payment of delinquent fines, but the

---

[6] To calculate this number, we note that the never-takers in the text groups pay an average of £3.06. Assuming that the compliers comprise 55.3 percent of the control group, we back out the fact that

Table 3. Results of the Phase 2 trial, by experimental condition.

Phase 2	N	Avg. amount paid (in pounds)	95% Confidence interval of average amount paid[a]	Percentage of subjects reached by text	Avg. amount paid (in pounds) by those actually reached	95% Confidence interval of average amount paid by those reached[b]
NONE	0					
STANDARD	912	8.34	[6.32, 10.55]	53.8	15.40	[10.76, 19.42]
AMOUNT	890	8.82	[6.51, 11.39]	55.7	15.83	[11.76, 20.34]
PERSONAL	917	11.21	[8.33, 14.43]	54.4	20.58	[15.40, 26.37]
PERSONAL/ AMOUNT	914	9.68	[7.34, 12.26]	52.6	18.34	[13.99, 23.12]

[a]Calculated using 100,000 bootstrap samples.

[b]Calculated using the package in R; for more on this procedure, see Gerber and Green (2012, Chapter 3).

best text message remains an open question. Our initial results suggest the following ordering:

PERSONAL > PERSONAL/AMOUNT > AMOUNT > STANDARD.

However, the amount of sampling variability surrounding each of the estimated treatment effects prevents us from ruling out the null hypothesis that all of the messages are equally effective.[7] We therefore conducted a second round of experiments in order to sharpen up our estimates of the effectiveness of each message type.

Phase 2 Trial

Unlike the design used in Phase 1, the Phase 2 trial does not include a NO TEXT group. In other respects, the design is the same, with each of the four message types assigned with equal probability. Table 3 shows the average amounts collected in fines from subjects assigned to each of the four experimental conditions. Table 3 also shows the rate at which text messages were successfully delivered to subjects. As in Phase 1, the proportion of compliers in Phase 2 is estimated to be slightly more than half of the 3,633 subjects (54.1 percent).[8] Interestingly, the relative ordering of the ITT and complier-average causal effect (CACE) estimates follows precisely the same pattern as in Phase 1. Given that there are 4! = 24 different ways of ordering the four message types, the odds that the same ordering would resurface by chance is just 0.042. The most effective treatment is again PERSONAL, which among compliers in Phase 2 generates on average of £5.20 more in fines than the STANDARD treatment.

untreated compliers in the control group paid an average of £5.59. The percentage gain associated with the CACE is computed by comparing the average payment rate of £17.31 among those in the treatment groups to this baseline payment rate among untreated compliers.

[7] Restricting our attention to those who actually received one of the four types of messages, we conducted an -test of equal CACEs and obtained a nonsignificant value of 0.424.

[8] Again, as expected, a chi-square test of equal proportions across the four conditions has a nonsignificant value of 0.61.

Table 4. Frequency distribution of amount paid by compliers, by experimental condition (both phases combined).

Amount paid in pounds	STANDARD	AMOUNT	PERSONAL	PERSONAL/AMOUNT
0	74.17	75.84	70.86	72.03
0.01 to 10	6.20	4.80	5.00	6.81
10.01 to 25	5.63	6.11	6.57	5.65
25.01 to 50	5.77	4.80	6.71	5.65
50.01 to 100	4.62	4.66	6.14	4.35
100.01 to 250	2.89	2.18	3.14	4.64
250.01+	0.72	1.60	1.57	0.87
Total	100.0	100.0	100.0	100.0
N	693	687	700	690
Average Payment	14.73	15.99	20.87	18.21

The two-tailed value comparing PERSONAL to STANDARD is 0.029, that comparing PERSONAL to AMOUNT is 0.097, and that comparing PERSONAL to PERSONAL/AMOUNT is 0.377.

Given the similarity in experimental design, context, and results between the two phases of the trial, it makes sense to pool the results to obtain the most precise sense of the relative performance of the four message types. Restricting the sample to compliers, Table 4 presents a frequency distribution depicting the amount paid across all four message types. The table reveals that the PERSONAL treatment produces both the highest rate of payments (29.1 percent) and the highest rate of payments in large amounts. Comparing PERSONAL to the STANDARD message, we obtain a two-tailed p value of 0.029 when testing the sharp null hypothesis that no subject became more likely to pay due to the change in message wording. The superiority of the PERSONAL treatment over the AMOUNT treatment is marginally significant using a two-tailed test *(p = 0.097)*. Less clear-cut is the superiority of PERSONAL over PERSONAL/AMOUNT. The fact that PERSONAL generates an average payment of £2.66 more than PERSONAL/AMOUNT is suggestive, but an absolute difference this large or larger would obtain by chance with approximately 0.38 probability even if the two treatments were equally effective. Although this test falls short of conventional 5 percent or 10 percent levels of statistical significance, the overall pattern of results implies that PERSONAL is the most effective message among the four we tested. Pooling both trials, the average fine paid by those who received the PERSONAL message is 41 percent greater than the average fine paid by those who received the STANDARD message, which was in use prior to the trial.

To put these findings into perspective (but without attempting a comprehensive benefit-cost analysis), it is helpful to calculate the expected increase in revenues associated with a shift in operating procedure. HMCTS handles approximately 500,000 cases per year. For our trial, mobile numbers were held for 23 percent of cases, and text messages were received by approximately 54 percent of people to whom they were sent. Suppose the baseline type of fine collection were no texting (the NONE treatment), and HMCTS were to adopt the STANDARD text. The ITT estimate of £4.16 from Table 2 implies that the short-term boost in revenues would be 500,000 x 0.23 x £4.16 = £478,400. Next, suppose that HMCTS were to replace the STANDARD text with the PERSONAL text. Table 4 indicates that the 62,100 people who would actually receive the text would on average pay £20.87 — £14.73 = £6.14 as a result of the PERSONAL treatment. This change in procedure would therefore boost total revenues by an additional 62,100 x £6.14 = £381,294. Taken together, these calculations indicate that the one-week boost in revenues of using personalized text message reminders would be approximately £860,000. This rough calculation ignores many additional considerations, such as the administrative savings from

not having to pursue debtors or the savings to debtors who would otherwise have to pay collection fees.

CONCLUSION

The current study may be interpreted narrowly as a pragmatic-randomized trial designed to improve the efficiency with which a government agency collects unpaid fines. In Phase 1, the trial demonstrated unambiguously that text messages from a judicial agency increased the amount that subjects paid in fines over the course of a week. Although only half of the subjects were reachable by text messaging (primarily because the court staff had not previously prioritized collecting valid mobile phone information after the fine was imposed by the court), texting tripled the amount that compliers paid in fines during the observation period. Comparing the effectiveness of alternative messages among those who received them, we see in both Phase 1 and Phase 2 indication that personalization was the message ingredient that most enhanced effectiveness. Personalization was effective on its own and, to a lesser extent, in combination with information indicating the amount owed. Our experiment does not furnish evidence about why personalization works best, but the success of the PERSONAL and PERSONAL/AMOUNT conditions provides an important first step in the development of theories about cognitive and social psychological mechanisms that cause people to comply with requests when addressed by name.

Whether these effects are sustainable over time remains an open empirical question. In principle, the novelty of receiving a text message from HMCTS could wear off as debtors become inured to this tactic. On the other hand, only 23.3 percent of the debtors in our trial had prior record of a distress warrant, which means that most recipients are likely to encounter this type of text message only once. As text messaging becomes a more common administrative tool for other governmental agencies, it remains to be seen whether the effects we observe here persist.

The larger message of this study is that randomized trials represent a feasible and cost-effective means of improving administrative efficiency. The creation and deployment of a randomized protocol demanded extra staff attention and effort, as did careful measurement of outcomes. Forbearance among debt collectors was required in order to hold out an untreated control group from Phase 1, the initial stage of our adaptive design. Nevertheless, these costs were more than offset by the value of information showing that messaging works and that previously untried messages work significantly better than the standard text messages. In the wake of this trial, HMCTS adopted the PERSONAL text treatment as part of its standard operating procedure.

LAURA C. HAYNES is Head of Policy Research, Behavioural Insights Team, and Visiting Research Fellow, Institute of Psychiatry, King's College London, Cabinet Office, 1 Horse Guards Road, London SW1A 2HQ, UK.

DONALD P. GREEN is Professor of Political Science at Columbia University, 420 West 118th Street, Mail Code 3320, New York, NY 10027.

RORY GALLAGHER is Senior Policy Advisor, Behavioural Insights Team, Cabinet Office, 1 Horse Guards Road, London SW1A 2HQ, UK.

PETER JOHN is Professor of Political Science and Public Policy, University College London, The Rubin Building, 29/30 Tavistock Square, London WC1H 9QU, UK.

DAVID TORGERSON is Professor, Department of Health Sciences, the University of York, York, Seebohm Rowntree Building, Heslington, York YO10 5DD, UK.

REFERENCES

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91, 444–455.

Aronow, P. M., & Samii, C. (2012). ri: R package for performing randomization-based inference for experiments. Version 0.9. [Software]. Retrieved from http://cran.r-project.org/web/packages/ri/index.html.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. Journal of Personality and Social Psychology, 43, 425–436.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. Journal of the Acoustical Society of America, 25, 975–979.

Dale, A., & Strauss, A. (2009). Don't forget to vote: Text message reminders as a mobilization tool. American Journal of Political Science, 53, 787–804.

Dijkstra, A. (2005). Working mechanisms of computer-tailored health education: Evidence from smoking cessation. Health Education Research, 20, 527–539.

Fellner, G., Sausgruber, R., & Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal, and social information. Journal of the European Economic Association, 11, 634–660.

Fjeldsoe, B., Marshall, A., & Miller, Y. (2009). Behavior change interventions delivered by mobile telephone short-message service. American Journal of Preventive Medicine, 36, 165–173.

Free, C., Knight, R., Robertson, S., Whittaker, R., Edwards, P., Zhou, W., Rodgers, A., Cairns, J., Kenward, M., & Roberts, I. (2011). Smoking cessation support delivered via mobile phone text messaging (Txt2stop): A single-blind, randomized trial. Lancet, 378, 49–55.

Gerber, A., & Green, D. (2012). Field experiments: Design, analysis, and interpretation. New York, NY: W.W. Norton.

Gerber, A., Green, D., Kaplan, E., & Kern, H. (2010). Baseline, placebo, and treatment: Efficient estimation for three-group experiments. Political Analysis, 18, 297–315.

Gleerup, M., Larsen, A., Leth-Petersen, S., & Togeby, M. (2010). The effect of feedback by text message (SMS) and email on household electricity consumption: Experimental evidence. The Energy Journal, 31, 113–132.

Iyer, G. S., Reckers, P. M. J., & Sanders, D. L. (2010). Increasing tax compliance in Washington state: A field experiment. National Tax Journal, 63, 7–32.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2010). Getting to the top of mind: How reminders increase saving. NBER Working Paper No. 16205. Cambridge, MA: National Bureau of Economic Research.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2010). Unwilling or unable to cheat? Evidence from a randomized tax audit experiment in Denmark. NBER Working Paper No. 15769. Cambridge, MA: National Bureau of Economic Research.

National Audit Office. (2007). Improving fines collection through better information on defendants' ability to pay. London, UK: Author. Retrieved from http://www.nao.org.uk/ wp-content/uploads/2007/06/fines_collection.pdf.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100, 322–331.

Shapiro, K., Caldwell, J., & Sorensen, R. (1997). Personal names and the attentional blink: A visual "cocktail party" effect. Journal of Experimental Psychology: Human Perception and Performance, 23, 504–514.

Slemrod, J., Blumenthal, M., & Christian, C. (2001). Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. Journal of Public Economics, 79, 455–483.

Sydes, M. R., Parmar, M. K., Mason, M. D., Clarke, N. W., Amos, C., Anderson, J., de Bono, J., Dearnaley, D. P., Dwyer, J., Green, C., Jovic, G., Ritchie, A. W., Russell, J. M., Sanders, K., Thalmann, G., & James, N. D. (2012). Flexible trial design in practice—Stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: A multi-arm multi-stage randomized controlled trial. Trials, 13, 168–181.

United Kingdom Ministry of Justice. (2011). Freedom of information request FOI-72392. Retrieved from Ministry of Justice website: http://www.justice.gov.uk/downloads/ information-access-rights/foi-disclosure-log/courts-tribunals/outstanding-court-fines.doc.

United Kingdom Ministry of Justice, Public Accounts Committee. (2012). Management of income. Retrieved from Parliament website: http://www.publications.parliament.uk/pa/ cm201012/cmselect/cmpubacc/1778/177806.htm.