DEADLINES AND MEMORY LIMITATIONS*

Steffen Altmann† Christian Traxler‡ Philipp Weinschenk§

August 2021

Forthcoming in Management Science

Abstract

This paper presents the results of two natural field experiments at a dental clinic. Guided by a simple theoretical model, we exogenously vary deadlines and associated rewards for arranging check-up appointments. Our data show strong and systematic effects of deadlines on patients' behavior. Imposing deadlines induces patients to act earlier and at a persistently higher frequency than without a deadline. We further document that individuals systematically respond to deadlines, even if these are not tied to explicit rewards. Several of our findings suggest that individuals' responses to deadlines are shaped by limitations in memory and attention. Our results illustrate that deadlines can be a powerful management tool to encourage timely task completion and to increase the cost effectiveness of performance-contingent rewards.

Keywords: Deadlines; memory limitations; limited attention; field experiment

JEL Codes: C93, D15, D91.

*An earlier version of this paper was titled 'Deadlines and Cognitive Limitations'. We thank Michele Belot, Keith M. Ericson, Kyle Hyndman, David Laibson, Dmitry Taubinsky, Klaus Wälde, and Joachim Winter for valuable comments and suggestions. The paper also benefited from fruitful discussions with seminar and conference participants in Bergamo, Berlin, Copenhagen, Düsseldorf, Frankfurt, Gothenburg, Harvard, Kiel, London, Mannheim, Mainz, Marburg, Munich, Prague, Rotterdam, and San Diego. Financial support from Volkswagen Foundation and the Max Planck Institute for Collective Goods is gratefully acknowledged. Ann-Kathrin Crede, Nicolas Kaufung, and Kyle Ott provided excellent research assistance.

†IZA Bonn and University of Copenhagen; altmann@iza.org

‡Hertie School; traxler@hertie-school.org

§University of Kaiserslautern; weinschenk@wiwi.uni-kl.de

1 Introduction

Deadlines are a pervasive feature of our lives. Companies and organizations use deadlines to manage workflows and steer their employees, e.g., by tying bonus payments to timely task completion. Consumers are offered time-limited coupons, rebates, and 'special deals'. After buying a product or subscribing to a service, consumer rights are governed by deadlines regarding refund and return policies, product warranties, (automatic) contract renewal, etc. In light of their prevalence in many important choice environments, systematic evidence on the behavioral effects of deadlines is surprisingly scarce.

This paper studies the impact of deadlines and variation in deadline length in two natural field experiments at a dental clinic. The task that people in our experiments face consists of calling the clinic to arrange a check-up appointment. Building upon a pre-existing system of postcard reminders, we introduce and randomly vary deadlines and associated economic incentives for making an appointment. In our main experiment, patients are eligible for receiving a reward if they contact the dentist within either one or three weeks after our intervention. A third group of patients faces no deadline. Orthogonal to the deadlines, we vary the level of rewards: patients can either receive a small dental care kit or a free professional dental cleaning worth approximately 70 euros. In addition, we study treatments that involve no explicit reward but simply ask the patients to contact the dentist before the communicated deadline. A smaller follow-up experiment tests, among others, the impact of longer deadlines of six and ten weeks.

The experimental design and our empirical analysis are guided by a simple theoretical framework, which analyzes how decision makers respond to deadlines for performing a one-off task. The model contrasts fully rational agents with agents who may turn inattentive and no longer have the task on top of their mind. The latter analysis is inspired by a growing body of literature indicating that individuals commonly exhibit limitations in memory and attention (e.g., Ericson, 2011; Karlan et al., 2016; Haushofer, 2015).[1] Paralleling earlier theoretical work by Taubinsky (2014) and Ericson (2017), our analysis shows that memory limitations can make a big difference for individuals' behavior under different deadlines. Longer deadlines, for instance, generally lead to an overall higher task performance among fully rational agents – both in terms of within-deadline task completion and overall performance rates. When agents' memory is limited, however, this may no longer hold: facing a long deadline or no deadline at all, these agents are becoming more and more likely to turn inattentive and never perform the task. As a result, performance rates of 'forgetful' agents may be higher under relatively short deadlines.

[1]Such cognitive limitations have been found to affect behavior in numerous contexts, including individuals' consumption decisions (Hossain and Morgan, 2006; Lacetera et al., 2012; Englmaier et al., 2018), responses to taxation (Chetty et al., 2009; Taubinsky and Rees-Jones, 2018), or the take-up of social benefits (Bhargava and Manoli, 2015).

To shed light on the empirical relevance of the differential predictions from our theoretical framework, we make use of data covering roughly 3,600 patient-treatment observations and rich information on patients' background characteristics. Our data reveal strong and systematic effects of deadlines on the timing and frequency at which patients arrange check-up appointments. In our main experiment, response rates after the first week of the intervention are significantly higher under a one-week than under a three-week deadline or in the condition without deadlines. After three weeks, cumulative response rates are similar under the one- and three-week deadline, with both values lying significantly above the response rate in the no-deadline treatment. The gap between the treatments with and without a deadline appears to be persistent: even after 100 days, the fraction of patients who have arranged a check-up appointment in the deadline treatments is around 10% higher than in the no-deadline condition. Hence, imposing a deadline leads to higher responses rates compared to an otherwise identical no-deadline condition. We also find that individuals systematically respond to deadlines that are not tied to (explicit) economic rewards. Moreover, the mere deadline effects are – at least in the short-run – quantitatively indistinguishable from the impact of relatively high-powered incentives.

The observed response rates under the one- and three-week deadline are consistent with both, the fully rational benchmark and a model with (moderate) memory limitations. The fact that response rates in the no-deadline condition remain below the values of the two deadline treatments in the long run, however, points to the relevance of memory limitations for patients' reactions to deadlines. Further support for the role of memory limitations comes from a detailed analysis of the timing of patients' responses. For all treatments in both experiments, hazard rates are either decreasing or U-shaped over time. In addition, for a given period before the deadline, hazard rates are negatively related to how much time has passed since a patient received a postcard. Both findings are inconsistent with the predictions from our baseline model, but they are in line with the memory-limitations framework. In principle, the patterns observed for population-level hazards could be driven by differential sorting of heterogeneous patient 'types' (see, e.g., Salant 1977 and, for a more recent discussion, Heffetz et al. 2021). However, a number of robustness checks – building on numerical simulations and complementary empirical approaches to construct more homogeneous subgroups of patients – suggest that unobserved heterogeneity is unlikely to be the main driver behind our findings on population hazards.

Our empirical results are corroborated in the follow-up experiment. The latter also provides evidence suggesting that relatively short deadlines can indeed trigger higher within-deadline response rates than longer deadlines. In particular, we find that the fraction of patients contacting the clinic before a three-week deadline is higher than the corresponding fraction that does so during the twice as long time window of a six-week deadline. While the difference is not statistically significant, it is consistent with the idea that short deadlines may not only accelerate but also boost overall task performance. More generally, the within-deadline response rates in the follow-up experiment are non-monotonic in the deadline length, suggesting that there might exist a 'sweet-spot' deadline length that maximizes timely task completion.

In addition to the field experiments, we also gathered data from a post-experimental survey at the clinic and a large, online survey experiment. The results from both surveys again underscore the potential benefits of relatively short deadlines: almost all survey participants indicate that they perceive deadlines as helpful to avoid problems related to 'postponing and forgetting' and a majority of respondents generally prefer relatively tight deadlines, suggesting that they are at least partially aware of the risk of not responding under a longer deadline.

The findings from our experiments provide new evidence on how deadlines and associated incentives affect individuals' completion of costly tasks. The existing empirical literature on deadlines has remained rather scattered and primarily comes from classroom, lab, or online experiments (Ariely and Wertenbroch, 2002; Burger et al., 2011; Bisin and Hyndman, 2020).[2] Perhaps most closely related to ours is the work by Shu and Gneezy (2010), who report evidence from surveys and two small-scale field experiments that study procrastination of enjoyable experiences in a student population. In line with our results, their evidence documents higher redemption rates of coupons with shorter redemption deadlines. We contribute to this literature in a number of important ways. First, we combine insights from a simple theoretical framework with evidence from large, natural field experiments that exogenously vary deadlines and associated incentives. Second, our data track patients' reactions to deadlines over a long time horizon. This allows us to study individuals' responses in the pre-deadline period as well as after having missed the deadline. Third, our setting pairs a long observation period with random variation in whether or not individuals face a deadline. The combination of these features allow us to provide direct causal evidence that imposing a deadline can lead to persistently higher task performance rates than an otherwise identical no-deadline environment.

By pointing to the role of memory limitations, our findings also contribute to the literature that theoretically analyzes how cognitive constraints (Taubinsky, 2014), present bias (O'Donoghue and Rabin, 1999; Herweg and Müller, 2011; Bisin and Hyndman, 2020), or the interplay between the two (Ericson, 2017) influence people's reactions to deadlines. While parts of our theoretical analysis are closely related to earlier work by Taubinsky (2014) and Ericson (2017), we complement their analysis by pointing to the role of hazard rates as an additional metric to theoretically analyze the interplay between deadlines, incentives, and memory

[2]A more indirectly related set of papers studies how time limits and ending rules in fundraising campaigns and online auctions affect charitable giving (Knowles and Servatka, 2015; Damgaard and Gravert, 2017) and strategic bidding (Roth and Ockenfels, 2002), respectively. The role of deadlines in enforcement problems is discussed in Dusek et al. (2020) and Heffetz et al. (2021).

limitations.[3] In addition, our paper adds new findings to the existing evidence reported in these studies. Based on a large-scale field setting, we document empirically how the length of deadlines and associated incentives shape individuals' task performance. When taken together with the evidence on cumulative task performance, our analysis of hazard rates and the complementary assessment of type heterogeneity suggest that memory limitations are indeed a relevant factor driving individuals' responses to deadlines.

Finally, our paper offers several insights for the academic analysis and practical use of deadlines in management. The corresponding literature has emphasized the potential of deadlines as an instrument to structure workflows, manage teams (Saez-Marti and Sjögren, 2008; Campbell et al., 2014; Weinschenk, 2016; Balasubramanian et al., 2018), or influence customer behavior (Bertrand et al., 2010; Shu and Gneezy, 2010). We contribute to this literature by documenting that deadlines can 'work', even if no reward is attached to them. This finding can potentially be attributed to non-pecuniary incentives and a signaling motive to meet the deadline, or to the role of deadlines as a means to better plan and structure one's tasks, akin to planning prompts (Rogers et al., 2015; Beshears et al., 2016). Furthermore, our results illustrate that deadlines can be a powerful management tool to increase the cost-effectiveness of performance-contingent rewards. As compared to open-ended incentives, rewards that are tied to a reasonably short deadline may induce early task completion without diminishing overall performance in the longer run. The contingency of the reward can thus accelerate performance, while reducing overall cost. Caution is warranted, however, as this result hinges on a sufficiently high post-deadline performance rate, which likely depends on the specific application considered.

In the following section, we introduce our theoretical framework and derive predictions for agents with limited and unlimited memory. Section 3 discusses the setup and procedures of our experiment. Section 4 presents our empirical results. Section 5 discusses our findings and Section 6 concludes.

2 Theoretical Framework

This section introduces a simple theoretical framework to analyze how deadlines and associated economic incentives influence an agent's decision to carry out a one-off task. To build intuition, we focus on a basic version of the model that deliberately abstracts from a number of features of our experimental setting. At the end of the section, we discuss a number of model extensions that illustrate how these features can be incorporated in our theoretical framework. A formal derivation of the main theoretical results and the analysis of the model extensions are presented in the Online Appendix.

[3] Our theoretical framework focuses on variation in the deadline length. In contrast, a key question in Ericson's (2017) analysis is how to set reminders and deadlines optimally to help present-biased agents with limited memory. Taubinsky (2014), in turn, focuses on how cues (e.g., reminders) and rehearsal affect attention dynamics and task completion of forgetful agents.

2.1 Rational agent with unlimited memory

An agent is confronted with a task that can be carried out at any time t before a deadline T. The costs of performing the task in period t are $c_t \in \mathbb{R}^+$. These reflect varying opportunity costs and are drawn each period from a distribution F. After observing the cost realization in period t, the agent decides whether to carry out the task in this period or not. If the agent completes the task, she obtains a reward y. If she does not carry out the task and the deadline has not yet been reached, the agent advances to the next period. The decision problem ends when the task is completed or when the deadline is reached. Unless stated differently, T is assumed to be finite.

Performing the task in period t yields a payoff of $y - c_t$, while not doing so yields δV_{t+1}, where δ is the discount factor and V_{t+1} the expected payoff – or option value – of reaching period $t+1$. Carrying out the task in period t is optimal for the agent if and only if the costs are sufficiently low:

$$y - c_t \geq \delta V_{t+1} \iff c_t \leq \hat{c}_t := y - \delta V_{t+1}. \tag{1}$$

The probability that the agent performs the task in period t, conditional on not having done so before – i.e., the *hazard rate* of carrying out the task in t – is thus

$$h_t = F(\hat{c}_t). \tag{2}$$

The probability that the agent completes the task until period t – the *cumulative performance rate* – is hence

$$p(t) = 1 - \Pi_{s=1}^{t}(1 - h_s), \tag{3}$$

where $\Pi_{s=1}^{t}(1 - h_s)$ measures the probability of no performance until t. The *within-deadline performance rate* is given by $p(T)$.

The option value is endogenous. Using the decision rule (1), the option value of reaching period t is given by

$$V_t = \int_0^{\hat{c}_t} (y - c)dF(c) + (1 - F(\hat{c}_t))\, \delta V_{t+1}. \tag{4}$$

The option value in t thus equals the probability of carrying out the task in t times the expected payoff of doing so plus the probability of not completing the task in t times the discounted option value of reaching $t+1$.

We obtain the following key properties of the option value. First, the option value is decreasing over time. Formally, $V_t > V_{t+1}$ for all $t \leq T$. Together with (1) this property directly implies that the cost threshold and therefore also the hazard rate is increasing over time. Second, due to the recursive structure, the option value –

and thus also the cost-threshold – depends on the distance to the deadline, $T - t$, but not on the absolute value of time t. We can then derive the following predictions (see Online Appendix A):[4]

PREDICTIONS 1: NO MEMORY LIMITATIONS

(i) The probability that the agent performs the task within the deadline increases in the length of the deadline. Formally, $p(T'') > p(T')$ holds for any T'' and T', where $T'' > T'$ and T'' could be finite or infinite.

(ii) The cumulative performance rate $p(t)$ in any period $t \leq T'$ is higher under a short deadline T' than under a longer deadline T'', where $T'' > T'$ and T'' could be finite or infinite. After the short deadline has passed, the difference in the cumulative performance rates decreases and eventually reverses, such that long-run performance rates are higher under the longer deadline than under the short deadline.

(iii) For a finite deadline, the hazard rate increases towards the deadline, $h_t < h_{t+1}$. The hazard rate depends on the distance to the deadline, but not on how much time has passed since the first period. For an infinitely long deadline, the hazard rate is constant over time, $h_t = h_{t+1}$.

The intuition behind these results is as follows. First, an agent who faces a longer deadline has more opportunities to draw sufficiently low costs to render task completion optimal. The within-deadline performance rate is hence increasing in the length of the deadline.

Second, relative to longer deadlines, short-run task performance is higher under short deadlines: tight deadlines induce the agent to optimally carry out the task in early periods – even for relatively high cost realizations. Over a longer time horizon, however, this benefit from short deadlines is overcompensated by the fact that long deadlines provide the agent with further opportunities to carry out the task in later periods. As compared to longer deadlines, short deadlines are thus more effective in terms of task performance within short time horizons, but less effective over longer time horizons.

Third, as time passes, the agent has fewer periods left until the deadline and therefore more pressure to perform the task. As the deadline comes closer, the agent becomes willing to carry out the task at higher cost realizations. Consequently, the hazard rate increases over time. Note further that the willingness to bear certain opportunity costs solely depends on how much time is left until the deadline. For a given time left until the deadline, the hazard rate is thus independent of the time that has passed since the first period. Finally, if the agent has infinitely many periods in which she can carry out the task – i.e., in case she faces no deadline – the hazard rate is constant over time.

[4] It is worth noting that these predictions (as well as the predictions for the case of memory limitations) do not require further assumptions on the reward, the discount rate, or the distribution of costs.

2.2 Agent with limited memory

We next consider an agent with limited memory, who may stop paying attention to the task. This is formalized in the simplest possible way: at the end of each period, the agent turns inattentive with probability $\gamma \in (0, 1]$. Once the task has dropped from the top of the agent's mind, she cannot perform the task anymore. In the basic version of the model, we further assume that the agent is naive about her memory limitations. We obtain the following predictions.

PREDICTIONS 2: MEMORY LIMITATIONS

(i) The probability that the agent performs the task within the deadline, $p(T)$, is increasing in T if γ is sufficiently low and decreasing in T if γ is sufficiently high.

(ii) The cumulative performance rate $p(t)$ in any period $t \leq T'$ is higher under a short deadline T' than under a long deadline T'', where $T'' > T'$ and T'' could be finite or infinite. If γ is sufficiently high, the same holds true for all periods $t > T'$.

(iii) For a finite deadline, the hazard rate is locally increasing, $h_t < h_{t+1}$, if γ is sufficiently low, whereas it is locally decreasing, $h_t > h_{t+1}$, if γ is sufficiently high. The hazard rate depends on the distance to the deadline and on the time since the first period. For an infinitely long deadline, the hazard rate decreases over time, $h_t > h_{t+1}$, for any $\gamma > 0$.

Part (i) of the prediction highlights that, in case the agent's memory is limited, within-deadline performance rates *can be higher* under shorter than under longer deadlines. Although a longer deadline provides additional opportunities to complete the task, it also softens the time pressure and thus induces the agent to carry out the task only for rather low cost realizations. If memory limitations are sufficiently strong, it is unlikely that the agent is able to use the additional opportunities provided by a longer deadline, since the task has likely dropped from her mind when reaching these later periods. Accordingly, a longer deadline may lower the probability that the agent performs the task within the deadline. This prediction echoes similar results on the impact of deadlines in the models of Taubinsky (2014) and Ericson (2017).

As in the case without memory limitations, an agent who faces a long deadline will carry out the task in early periods only for relatively low cost realizations. Thus, long deadlines lead to lower short-run performance rates than shorter deadlines. If the agent's memory limitations are sufficiently strong, however, there will be no overtaking of the cumulative performance rates in later periods. Short deadlines may thus not only increase cumulative performance rates in the short run but give rise to persistently higher performance rates than longer (or infinitely long) deadlines. It is worth emphasizing that the first two parts of the prediction imply that, for sufficiently strong memory limitations, short deadlines both accelerate and increase overall task performance. Hence, with memory limitations, there is scope for a double dividend of imposing short deadlines.

Part (iii) of the prediction reflects two countervailing effects. To see this, note that the hazard ratio between two consecutive periods can be expressed as

$$\frac{h_{t+1}}{h_t} = \underbrace{\frac{F(\hat{c}_{t+1})}{F(\hat{c}_t)}}_{\text{(A) Option-value effect}} \underbrace{(1-\gamma)(1-F(\hat{c}_t))\frac{1-\sum_{s=1}^{t-1} F(\hat{c}_s)(1-\gamma)^{s-1}\prod_{r=1}^{s-1}(1-F(\hat{c}_r))}{1-\sum_{s=1}^{t} F(\hat{c}_s)(1-\gamma)^{s-1}\prod_{r=1}^{s-1}(1-F(\hat{c}_r))}}_{\text{(B) Memory-limitation effect}}. \tag{5}$$

Term A captures the option-value effect: since the option value is decreasing over time, the cost threshold is higher in period $t+1$ than in period t, which contributes towards an increasing hazard rate over time (Term A > 1). Term B captures the memory-limitation effect: due to limited memory, the likelihood that the agent has the task on the top of her mind is decreasing over time, which works towards a decreasing hazard rate ($0 \leq$ Term B < 1). Depending on the relative strength of the two effects, the hazard rate can thus be locally increasing or decreasing. Notably, the option-value effect is relatively strong in periods close to the deadline, whereas it is relatively weak in early periods where the deadline is still far away (see Online Appendix A). The hazard rate might therefore exhibit a U-shaped pattern over time. The memory-limitation effect further implies that the hazard rate does not only depend on how much time is left until the deadline, but also on how much time has passed since the first period. For an agent who faces no deadline, the memory-limitation effect always dominates and the hazard rate thus decreases over time.

Table 1 summarizes key predictions for an agent with and without memory limitations, respectively. The table focuses on outcomes for which the model variants may yield differential predictions.

Table 1: Summary of Theoretical Predictions

Outcome	Variable	Agent with unlimited memory	Agent with memory limitations
Within-deadline performance rate	$p(T)$	increasing in T	increasing in T if γ is low, decreasing in T if γ is high
Cumulative performance rate	$p(t)$	higher in the long-run for long T	higher in the long-run for short T if γ is high
Hazard rate	h_t	monotonically increasing; only depends on $T-t$	locally decreasing for sufficiently high γ; depends on $T-t$ and t

Notes: The table presents key predictions of our basic theoretical framework. T is the deadline length; t denotes the period (time); γ is the probability that the agent turns inattentive in a given period.

2.3 Model extensions and relation to experiments

In the Online Appendix, we discuss a variety of extensions to our basic theoretical framework. First, we examine the case in which the agent is sophisticated about her memory limitations. We show that – in anticipation of her possible future inattention – a sophisticated agent is willing to bear higher costs of task performance than a naive agent. Importantly, however, all key qualitative differences between Predictions 1 and 2 also prevail for the case in which the agent is sophisticated about her memory limitations. Second, we allow for present bias and overoptimism as alternative factors that might influence an agent's response to deadlines. Our analysis shows that these factors alter the cost threshold in (1), but otherwise do not change the main qualitative results from our baseline model without memory limitations.

In a further set of extensions, we adapt our general theoretical framework by incorporating features that are specific to our experimental setting: 1) We allow for positive post-deadline rewards and task completion after the deadline. This is important in our empirical setting, in which participants can (and do) arrange check-up appointments after the deadline has passed. A natural driver of such post-deadline responses in our experiment is that patients of the dental clinic face further incentives to arrange a check-up appointment, beyond the explicit rewards that are tied to the deadlines.[5] 2) Participants in our experiment face costs for, both, arranging as well as attending check-ups. The model can readily be reinterpreted to capture both of these costs.

Beyond these extensions, two further points are worth stressing. First, patients in our experiment might have non-pecuniary incentives to comply with the deadline: they could exhibit social image concerns (e.g., call in a timely manner to signal to the dentist that they care about their dental health) or perceive the deadline as a medical recommendation. Accordingly, one should interpret the reward y in broad terms, i.e., as the sum of all explicit and implicit rewards that accrue when the agent complies with the deadline. Notably, the presence of implicit incentives to comply with a deadline render Predictions 1 and 2 from above also relevant for treatments in which no explicit rewards are provided. We will return to this point in Sections 4 and 5 below. Finally, note that our theoretical analysis derives individual-level predictions. Our empirical results, in contrast, will mainly be based on population-level observations. This distinction is particularly relevant for the analysis of hazard rates. Independent of the channels discussed above, differential sorting of heterogeneous types could alter the shape of population-level hazard rates over time. We will address this point in Section 4.2, where we analyze the case of type heterogeneity in our context.

[5] These may stem from implicit incentives related to (expected) improvements in future health, or explicit economic incentives to regularly arrange check-ups that are independent of our intervention. Participants covered by the German public health insurance (more than 80% of our sample), for instance, receive a 20% [30%] allowance on treatment costs if they had at least one check-up per year for the last 5 [10] years.

3 The Experiments

We study the effects of deadlines on task performance in two field experiments at a large dental clinic. The experiments exogenously vary deadlines and associated economic incentives to arrange appointments for preventive check-ups. We examine how deadlines affect the timing and frequency at which patients arrange appointments. Several features of our setting make it ideally suited to explore the interplay between deadlines, task performance, and memory limitations. Check-ups should be conducted regularly, but at a relatively low frequency.[6] Patients at our collaboration partner's clinic have to actively contact the dentist to arrange a new check-up appointment. Hence, we study a setting in which people have to actively carry out a task, but have no well-defined date for when to do so. This is similar to a variety of other tasks, like adapting one's insurance or savings plans, switching energy or phone providers, etc. (e.g., Chetty et al. 2014, Heiss et al. 2016).

Anticipating that patients frequently fail to regularly schedule check-up appointments,[7] the dentist uses a reminder system. Whenever a patient is up for the next check-up according to her recommended check-up interval (see fn. 6), the patient receives a postcard reminding her to arrange an appointment. Making use of this pre-existing reminder system, our experiments introduce and exogenously vary deadlines and incentives for arranging appointments. If a patient contacts the dentist until a specified deadline, she is eligible to receive a reward (see below). Note that after receiving a postcard, patients have the task on the top of their mind – just as in the first period of our theoretical framework. The costs of subsequently carrying out the task, however, may fluctuate, for example due to other chores or duties. Over time, the task might thus drop from the top of a patient's mind if her memory is limited.

3.1 Main Experiment

In our main experiment, we varied the length of the deadline across three different treatment conditions. In the *D1* treatment, patients faced a short deadline that was set one week after the day at which postcards were sent. In treatment *D3*, we implemented a longer deadline of three weeks. Finally, in the *ND* treatment, the postcards contained no deadline; the reward was thus available for an unlimited time period.

[6] In the case of dental health, it is widely agreed that regular check-ups are a key factor of effective health prevention, since many dental diseases are asymptomatic in their early stages but can be discovered and treated through professional check-ups (see, e.g., Lang et al. 1994). Dentists in Germany typically suggest check-up intervals of 6 months and somewhat shorter intervals of 3–4 months for patients with an elevated risk of developing certain dental diseases.

[7] In an independent online survey that we conducted with roughly 3,000 individuals, 56% of respondents indicate that they aim at having (at least) two dental check-ups per year. Among this group, 43% did not act accordingly during the past 2 years. Compare Section 5 and Section C.5 in the Online Appendix for further details.

We implemented two different reward levels: a large and a small reward. The large reward consisted of a voucher for a free professional dental cleaning, which would otherwise cost about 70 euros.[8] In the small reward condition, patients could choose between various dental care products, such as toothpaste, dental floss, etc. In addition, we conducted treatments that involved no explicit reward, but simply asked the patients to contact the dentist before the communicated deadline. Interacting the three deadline treatments (D1, D3 and ND) with the three reward levels leaves us with a total of 9 treatment cells. Table 2 illustrates the 3×3 factorial design. (An example of a reminder postcard and more detailed information of the mailing procedures are provided in the Online Appendix.) Since we are primarily interested in the impact of deadlines, we will focus on the deadline dimension when discussing our results in Section 4. Most of our analysis therefore pools observations with identical deadline lengths that involve different rewards. This approach is valid as we assign people randomly and independently to the different treatment cells. To compare the impact of deadlines to the effect of differences in economic rewards, however, we also evaluate the effects of the different incentives.

Table 2: Treatments of the Main Experiment

	D1 1-week deadline	**D3** 3-week deadline	**ND** No deadline
Large reward	D1-LR $N = 301$	D3-LR $N = 291$	ND-LR $N = 294$
Small reward	D1-SR $N = 301$	D3-SR $N = 295$	ND-SR $N = 291$
No explicit reward	D1-NR $N = 305$	D3-NR $N = 293$	ND-NR $N = 290$

Notes: The main experiment comprised N = 2,661 patient-treatment observations (covering 1,175 individuals from 1,015 households). Summary statistics of patients' background characteristics and randomization checks are presented in Table A.1 in the appendix.

The main experiment was conducted between 2011 and 2013. In 43 biweekly sending waves w, we randomly assigned patients who were up for their next check-up to treatments.[9] To minimize risks associated with treatment spillovers, we assigned members of the same household to the same treatment (within each randomization wave). In total, we observe N = 2,661 patient-treatment observations, covering 1,175 individuals from 1,015 households. These numbers reflect that, as a result of the long overall duration of the experiment and the

[8] One might wonder whether patients perceive a professional dental cleaning as a reward. Reassuringly, 64% of the participants in our online survey (see fn. 7) state that they do not find the procedure to be unpleasant; 93% would 'rather' or 'definitely' make use of the procedure if it was offered free of charge by their health insurance plan. Regarding the smaller reward, 73% of respondents in our online survey (correctly) expect a 'small present' in this context to be worth roughly 2.50 euros (see Table XI in the Online Appendix).

[9] Some patients might 'select out' of the experiment by already arranging a check-up well before their next due date. Our data suggests that this selection effect is, at best, modest.

recommended check-up interval of typically six months, most patients were treated more than once. Specifically, 364 (359) [291] individuals in our sample received one (two) [three] postcards throughout the experiment; 161 were treated more than 3 times. Our procedure ensures that treatments are randomized independently, each time a patient is up for a new appointment.[10]

We obtained rich individual-level data, which include sociodemographic as well as health-related variables, such as patients' health insurance status, history of major dental treatments, and their dental health classification ('at risk'). Summary statistics and balancing checks indicate that randomization succeeded in generating samples that are well balanced across treatments (see Table A.1 in the Appendix).

Our main dependent variable is the date at which a patient contacts the dentist to arrange a check-up appointment. We derive this variable from a data set that covers all inbound phone calls at the dental clinic between 2011 and mid 2014. As indicated above, the date at which a participant first contacts the dentist to make the appointment (rather than the actual date of the appointment) is the relevant date which determines whether she is eligible for receiving the reward. It is also a natural outcome variable, since it is the first observable response of patients. Finally, using this response measure instead of the date of the check-up avoids potential issues of congestion in the dentist's schedule. We nevertheless do measure whether a participant missed (or re-scheduled) an appointment. Such 'no-shows' are rare in our sample: the average no-show rate is 4.6% and does not systematically vary across treatments.

The empirical analysis will focus on patients' responses during the first 100 days after the sending date of each wave. Studying longer response periods is potentially problematic, because some individuals (e.g., diabetes mellitus patients and other patient groups who exhibit an increased risk of developing certain dental diseases) will already receive their next check-up reminder after 3–4 months.

3.2 Follow-up Experiment

During the 2nd and 3rd quarter of 2013, we conducted a second experiment in the same institutional setting. Besides assessing the general robustness of our main results, we wanted to explore whether our key findings are qualitatively stable under longer deadlines. In addition to treatments paralleling D1, D3, and ND from the main experiment, the follow-up experiment thus included treatments with a 6-week (*D6*), a 10-week (*D10*), and an 'end-of-year' deadline set to December 31 (*EoY*). As a reward, the follow-up experiment involved a dental-care kit worth 10 euros. In all other respects, the procedures were identical to those in our main experiment.

[10]The pre-existing reminder system at the clinic is also ideal from a methodological perspective, as it provides a natural environment that is unlikely to raise questions about experimenter scrutiny or an ongoing experiment: patients of this clinic are already used to (a) receiving reminder messages, and (b) to variation in these reminders, as the dentist regularly modified the design and content of reminder postcards in the past (see Altmann and Traxler 2014).

Table 3: Treatments of the Follow-Up Experiment

Treatments:	D1	D3	D6	D10	EoY	ND
Deadline:	1 week	3 weeks	6 weeks	10 weeks	$\approx$ 6 months	–
N	169	163	163	166	91	175

Notes: Depending on the randomization wave, the 'end-of-year' (*EoY*) deadline corresponded to a deadline length of 5–7 months. Due to a delayed roll out, the number of observations is slightly smaller in the *EoY* treatment. The overall number of patient-treatment observations is $N = 927$ (covering 798 individuals from 642 households). Table A.2 in the appendix presents summary statistics of patients' background characteristics and randomization checks.

Table 3 summarizes the different treatments (an overview of participant characteristics and balancing tests for the follow-up experiment are provided in Table A.2 in the Appendix). The higher number of deadlines and the lower overall number of $N = 927$ patient-treatment observations implies that treatment cells in the follow-up experiment are smaller than the ones from our main experiment. Consequently, the statistical power in our analyses of the follow-up experiment is somewhat limited.[11] Hence, the second experiment should primarily be seen as a qualitative replication and extension of our main experiment.

4 Empirical Results

The discussion of our empirical results will be structured according to the outcomes and main qualitative predictions derived in our theoretical framework in Section 2. To evaluate the behavioral consequences of deadlines, we will first focus on (a) *cumulative performance rates* until a given point in time t and (b) the resulting *within-deadline performance rates*. Based on the exact date at which patients contact the dentist to arrange a check-up appointment, one can readily derive empirical measures for these outcomes. In particular, we compute *cumulative response rates* (i.e., the fraction of patients who have contacted the dentist within t days after the intervention) as well as *within-deadline response rates* (the fraction of patients who have called before the communicated deadline).

To further analyze how limitations in memory and attention may mediate the observed deadline effects, we will also examine (c) the shape of *hazard rates* over time (i.e., the share of patients who respond on a given day t, conditional on not having done so before) and whether the observed hazard rates depend only on the time left until the deadline or also on the time passed since receiving the postcard. Note, however, that it is difficult to directly relate the empirically observed population hazards to our theoretical framework. If patients differ

[11] With a sample size of $\sim$165 per treatment and a distribution of responses similar to the one from the ND condition, the minimum detectable effect size (MDES) for response-rate differences in the follow-up experiment is about 10–15pp. This compares to a MDES of 4–6pp in the main experiment.

fundamentally – say, in their opportunity costs of arranging check-ups – they may be more or less responsive at different points in time. Differential sorting could then alter the shape of hazard rates over time (see, e.g., Salant 1977 and, for a more recent discussion, Heffetz et al. 2021). Our subsequent empirical analysis therefore focuses primarily on cumulative response rates (i.e., outcome (a) and (b)), where this concern does not apply. In Section 4.2, which studies hazard rates, we carefully examine the case of type heterogeneity in our context.

4.1 Response Rates

Main Experiment. In a first step, we study treatment differences in cumulative response rates in our main experiment. Figure 1 presents the evolution of response rates of patients facing a one-week deadline (light blue line), a three-week deadline (dark blue line), or no deadline (dashed grey line). To complement the graph, Table 4 presents linear-probability estimates of the following structure:

$$Y_{iw}^{t} = \beta_0 + \beta_1 \, \mathrm{D1}_{iw} + \beta_2 \, \mathrm{D3}_{iw} + X_{iw} \, \gamma + T_w \, \delta + \varepsilon_{iw}. \tag{6}$$

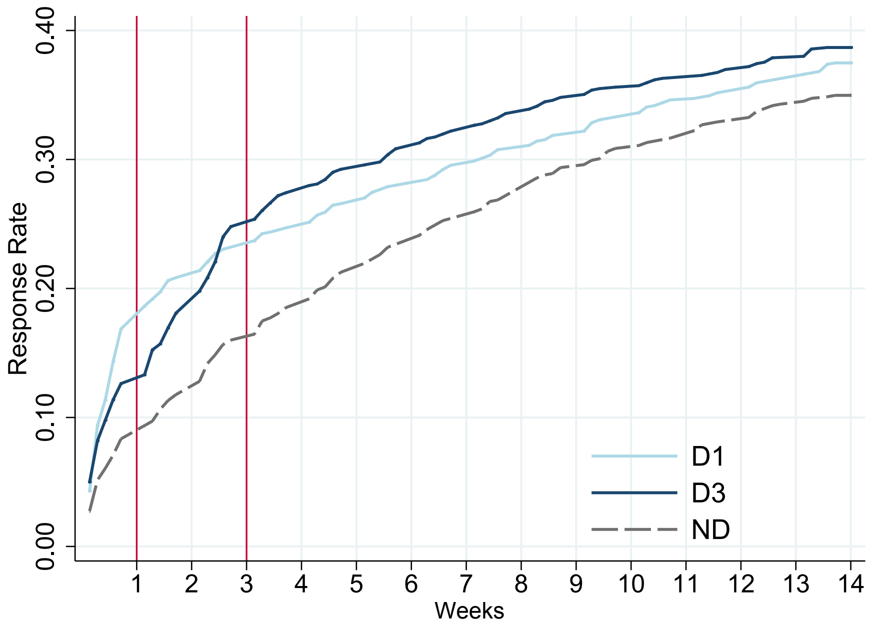
Y_{iw}^{t} indicates whether patient i called within t days after a reminder from sending wave w; D1 and D3 are treatment dummies, X and T are vectors of control variables and a full set of sending-wave dummies, respectively.

Figure 1 shows that, during the first week of the intervention, response rates under the one-week deadline lie strictly above the treatments with longer deadlines. At the end of the first week, the response rate in the D1 treatment is roughly 4pp (percentage points) higher than in the D3 treatment. This difference is statistically significant ($p = 0.017$ and $p = 0.011$ in specifications with and without controls, respectively; see the post-estimation tests for Columns 1 and 2 of Table 4). Compared to the no-deadline treatment, the one-week response rate is more than twice as high in the D1 treatment (16.8% vs. 8.3%, $p < 0.001$; see Columns 1 and 2), and the response rate in the D3 treatment still lies about 4pp above the rate in the ND treatment ($p < 0.001$). These findings underline the strong impact of imposing relatively tight deadlines of one or three weeks on early responses.

During the second and third week, the response rate in the D3 treatment catches up and, after about 2.5 weeks, surpasses the corresponding rate of the D1 treatment. At the end of the third week, the response rate in the D3 treatment is about 2pp higher than in the D1 treatment ($p = 0.401$ and $p = 0.352$; see Columns 3 and 4). For both deadline treatments, the three-week response rates are substantially higher than the rate observed in the no-deadline condition: the gap ranges from 7pp (D1, $p < 0.001$) to almost 9pp (D3, $p < 0.001$), corresponding to an increase in responses of 45–55% relative to the ND treatment.

After the third week, the ND treatment catches up a bit relative to the D1 and D3 treatment. However, the response rates of both deadline treatments remain well above those in the ND treatment. Even after 100

Figure 1: Cumulative Response Rates (Main Experiment)



Notes: The figure presents the empirically observed cumulative response rates over 14 weeks. *D1* and *D3* denote treatments with a deadline length of 1 and 3 weeks, respectively. *ND* denotes the treatment without deadline.

days, the response rate in the D3 [D1] treatment is 4.8pp [2.8pp] above the rate in the ND treatment ($p = 0.070$ [$p = 0.224$] when comparing D3 [D1] with ND, and $p = 0.082$ when jointly comparing the treatments with a deadline to the no-deadline case; see Columns 5–8). Complementary evidence from duration models corroborate the positive impact of imposing deadlines on patients' responses. Estimates from Cox proportional hazards models document roughly 15–20% higher (conditional) response rates in the D1 and D3 treatment, relative to the no-deadline case (see Table I in the Online Appendix).

The observation that response rates under both deadlines remain persistently above the no-deadline treatment highlights that imposing deadlines does not only induce early action, but may also positively affect task performance in the longer run. Beyond its practical importance, this finding also points to the potential relevance of memory limitations.[12] At the same time, the comparison of within-deadline response rates in the D1 and D3 treatment shows that the fraction of patients who respond within three weeks in the D3 treatment (24.8%) is significantly higher than the one-week response rate in D1 (16.9%; t-test, $p < 0.001$). This finding

[12]Remember from Section 2 that – absent memory limitations – cumulative response rates in the no-deadline treatment should catch-up and eventually surpass response rates for the deadline treatments. In principle, this catch-up process might be very slow and patients could redeem their vouchers much later. According to the dental clinic, however, this essentially never happened. In line with this anecdote, we continue to observe lower response rates in the ND treatment over an observation period of up to six months after the intervention. Recall, however, that the interpretation of such long-run outcomes is complicated by the fact that some participants are already due for the next check-up reminder.

Table 4: Treatment Effects on Response Rates (Main Experiment)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	7 Days		21 Days		100 Days		100 Days	
D1	0.085***	0.086***	0.070***	0.069***	0.028	0.028		
	(0.017)	(0.017)	(0.020)	(0.019)	(0.023)	(0.022)		
D3	0.043***	0.044***	0.088***	0.088***	0.042*	0.042*		
	(0.015)	(0.014)	(0.019)	(0.019)	(0.023)	(0.023)		
D1 or D3							0.035*	0.036*
							(0.020)	(0.020)
Large Reward (LR)		0.087***		0.090***		0.097***		0.096***
		(0.019)		(0.022)		(0.026)		(0.026)
Small Reward (SR)		-0.011		-0.022		0.028		0.025
		(0.014)		(0.018)		(0.022)		(0.022)
Constant	0.083***	0.020	0.160***	0.007	0.353***	-0.042	0.353***	0.028
	(0.010)	(0.074)	(0.014)	(0.097)	(0.019)	(0.121)	(0.019)	(0.119)
Post-Estimation Tests: (p-value)								
D1 = D3	0.017	0.011	0.401	0.352	0.564	0.502		
LR = SR		0.000		0.000		0.008		0.007
LR = D1		0.978		0.489		0.044		
LR = D3		0.066		0.932		0.110		
Controls	–	Yes	–	Yes	–	Yes	–	Yes

Notes: The table presents LPM estimates of equation (6). The coefficients capture the treatment effects on the cumulative response rate within 7, 21, and 100 days (N = 2,661). The lower part reports the p-values from Wald tests. LR and SR capture the Large and Small Reward treatments, respectively. Every second specification includes individual control variables and dummies absorbing wave-specific effects. Robust standard errors, clustered at household level (1,015 cluster), are reported in parentheses. ***, **, * indicates significance at the 1%-, 5%-, 10%-level, respectively.

is consistent with both, the fully rational baseline model and with (moderate) memory limitations (see Section 2). The finding of higher within-deadline response rates in D3 than in D1 also raises the question whether the observation of lower long-run response rates in the ND treatment are specific to the no-deadline case, i.e., whether facing no deadline at all is perceptually an extreme case that is not comparable to very long, but explicitly specified finite deadlines. We will return to this point when we discuss the results from the follow-up experiment.

Incentives. So far, we have documented that deadlines have a strong impact on the timing and overall rate at which participants respond to our intervention. In a next step, we examine the impact of economic incentives on patients' responses and compare them to the deadline effects documented above. The estimates from Table 4 show no statistically and economically significant differences between treatments involving a small reward and the no-incentive baseline. In contrast, the large reward strongly and persistently increases responses. Relative to treatments with no explicit reward, response rates in the large-reward treatments increase by almost 10pp.

This holds for all time spans considered in Table 4.[13] Comparing the large-reward coefficients to the estimated deadline effects from Table 4, we find that offering a large reward has a qualitatively and quantitatively similar effect on patients' one-week response rate as imposing a one-week deadline. Similarly, communicating a three-week deadline has an almost equally sized impact on response rates within the first three weeks as the large reward. Considering cumulative response rates after 100 days, the effect of the large reward tends to be stronger than the impact of either of the two deadlines, although the difference is statistically significant only relative to the D1 ($p = 0.044$), but not the D3 treatment ($p = 0.110$).

Next, we separately examine all nine treatment cells from our 3×3-design. The top, middle, and bottom panel of Figure 2 depict response rates for treatments in which individuals face large, small, or no explicit rewards, respectively. In line with the estimates discussed above, we observe that response rates are highest in the large-reward treatments. It is also worth noting that the persistent gap in response rates between the deadline treatments and the no-deadline condition is observed, both, in the presence of large rewards (top panel of Figure 2) and without explicit rewards (bottom panel). This highlights that the sluggish response in the no-deadline case is not a small-reward phenomenon. If anything, the longer-term gap in response rates is most pronounced in the treatments featuring large rewards. This finding suggests that individuals in the ND treatment may indeed have forgotten to redeem their voucher, rather than not finding it valuable enough to do so.

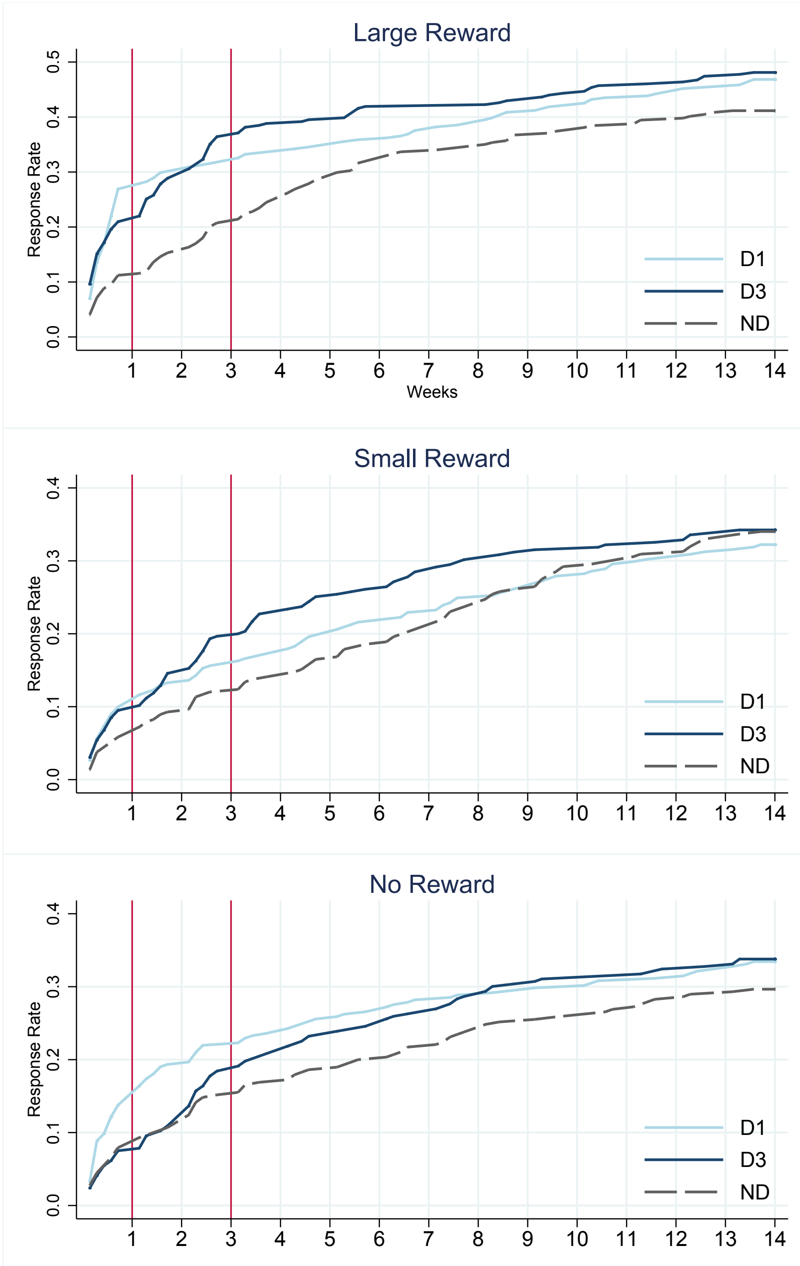
Finally, the results depicted in the bottom panel of Figure 2 indicate that deadlines affect behavior even when there is no explicit reward attached to them. From a practical perspective, the most simple and inexpensive treatment ('please call us within ... weeks') thus performs fairly well. There are several possible explanations for this finding. First, deadlines that are not tied to an explicit reward might be perceived as an informative signal of the (implicit) health rewards related to a check-up (e.g., signaling the dentist's perception of a patient's health status).[14] Second, as discussed in Section 2, patients might have non-pecuniary incentives to comply with the deadline (related to, e.g., social image concerns). Finally, deadlines in the no-reward treatments might function as a planning prompt (Rogers et al., 2015; Beshears et al., 2016), thereby amplifying the pure reminder effect from the postcards.

Follow-up Experiment. The persistently lower response rates in the no-deadline condition raise the question whether having no deadline at all is a special case or whether – consistent with the memory-limitations interpretation – longer deadlines can indeed trigger lower responses than shorter ones. We address this question in our

[13]The strong effect of the large reward is also documented in the duration analysis (see Table I in the Online Appendix).

[14]The evidence from our survey experiment, which is discussed in more detail in Section 5, is inconclusive on this point. After being exposed to a no-reward scenario, survey participants who face a deadline are more likely to state that the postcard may indicate a potential health problem (as compared to those in the corresponding ND scenario; see Table X in the Online Appendix). The difference is, however, not statistically significant.

Figure 2: Cumulative Response Rates for all 3×3 Treatment Cells of the Main Experiment



Notes: The figure presents the (empirically observed = raw) cumulative response rate over 14 weeks for each of the nine combinations of deadlines and rewards from our main experiment. *D1* and *D3* denote treatments with a deadline length of 1 and 3 weeks, respectively. *ND* denotes the treatment without deadline.

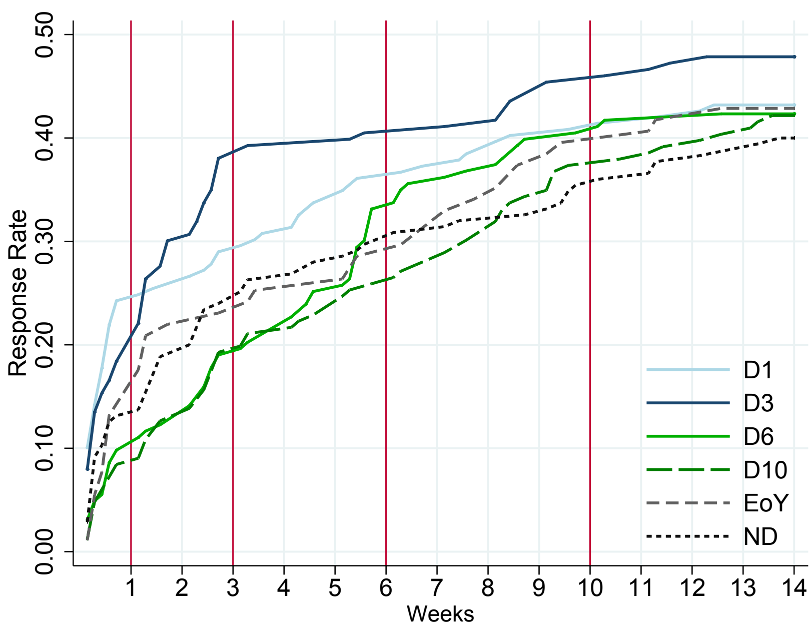
follow-up experiment by exposing participants to explicitly specified deadlines ranging from one week (D1) to ten weeks (D10 treatment). The cumulative response rates displayed in Figure 3 reveal that responses during the first three weeks follows a similar pattern as in the main experiment:[15] Responses in the D1 treatment surge

[15]The observed differences in response *levels* are hard to interpret: the follow-up experiment differs from our main experiment in the reward level, the sample, and the implementation period.

during the first week. During the following two weeks, the response rate in the D3 treatment catches up and overtakes the D1 treatment before the end of the three-week deadline (see the dark and light blue solid lines in Figure 3). After the end of the third week, response rates for the D1 and D3 treatments are substantially higher than those observed under longer (or no) deadlines (see Table IV in the Online Appendix for the corresponding LPM estimates).

In weeks 4–6, the response rate in the D6 treatment (solid light green line) increases relative to the D1 and D3 treatment. However, this catch-up is incomplete such that response rates after week six are still higher in D1 and D3 as compared to the D6 treatment. A similar pattern is observed during weeks 7 to 10: the response rate in the D10 treatment (dashed dark green line) catches up but remains well below the levels observed for treatments with shorter deadlines (i.e., D1, D3, and D6). After 100 days, response rates in the D1, D6, and D10 treatment are virtually identical, but the response rate in the D3 treatment remains roughly 6–9pp above the values in the treatments involving longer or no deadlines.

Figure 3: Cumulative Response Rates (Follow-up Experiment)



Notes: The figure presents the empirically observed cumulative response rates over 14 weeks. *D1, D3, D6, D10* denote treatments with a deadline length of 1, 3, 6, and 10 weeks, respectively. *ND* denotes the treatment without deadline. *EoY* denotes the treatment with an 'end-of-year' deadline (set to December 31).

Most notably, Figure 3 indicates that the rate at which individuals in the D3 treatment respond within the three-weeks deadline is higher than the 6-weeks response rate in the D6 treatment (38.0% vs 33.1%) and also marginally higher than the corresponding 10-weeks rate in D10 (37.3%). Hence, the fraction of patients who respond within the three-week deadline of the D3 treatment is roughly 5pp higher than the corresponding fraction of patients who respond within the twice as long time window in the D6 treatment. While the differences are not statistically significant, the observed response patterns speak against the notion that within-deadline response rates increase monotonically in the deadline length. In line with the predictions of our memory-limitations framework, the evidence suggests that longer deadlines may indeed diminish the likelihood of performing a task within the deadline.[16]

In sum, the follow-up experiment corroborates the key findings from our main experiment. The evidence illustrates that the persistently low response in the ND treatment of our main experiment is not a special case, but rather a limit case of facing longer and longer explicitly specified deadlines. The follow-up experiment further indicates that the fraction of people completing the task within the deadline can be higher under shorter deadlines. Overall, the evidence suggests that relatively tight deadlines can accelerate and boost task performance.

4.2 Hazard Rates

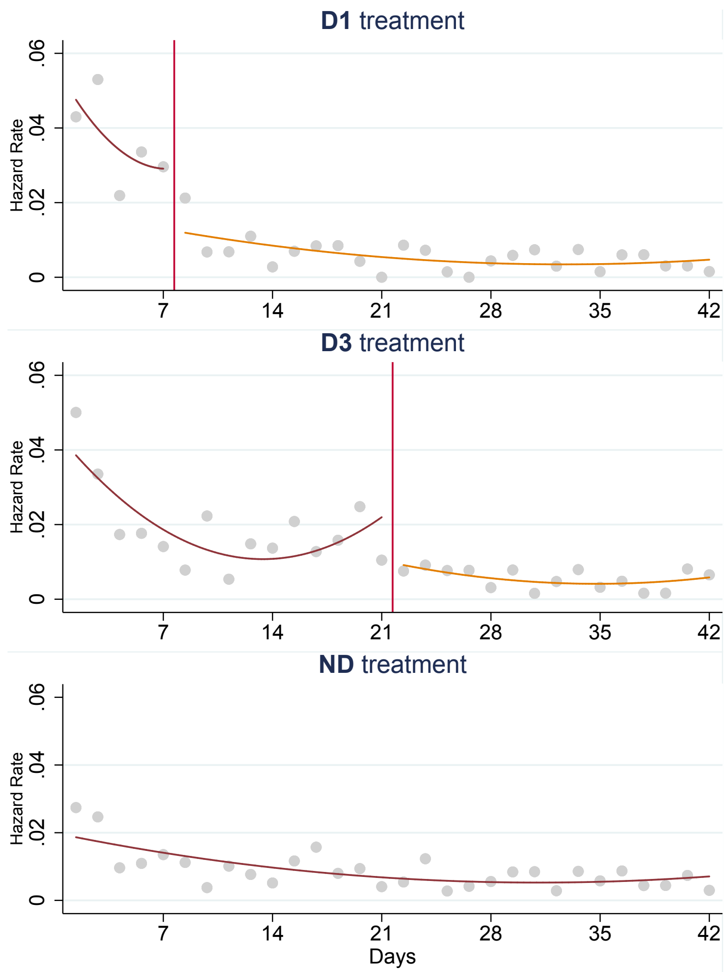
In a next step, we shift the focus from cumulative response rates to the exact *timing* of individuals' responses. In particular, we examine treatment differences in the shape of hazard rates – patients' likelihood to respond on a given day t, conditional on not having done so before. Recall that the shape of hazard rates over time might be influenced by heterogeneity across different patient types. We will closely examine this concern below.

Main Experiment. Figure 4 plots hazard rates for the D1, D3, and the ND treatment from our main experiment. The figure depicts the empirically observed hazard rates on a daily basis (gray dots) together with quadratic fits, estimated separately for the pre- and the post-deadline period. The fitted curves in the three panels of Figure 4 indicate that hazard rates in all treatments are decreasing initially. For the D1 treatment (top panel of Figure 4), this declining pattern is observed for the entire period until the deadline; for the no-deadline treatment (bottom panel), we see a declining pattern throughout our entire observation window. In the D3 treatment (middle panel), we first observe a decline in hazards, followed by a modest increase when moving from the second to the third week. Over the three weeks before the deadline, the hazard rate in the D3 treatment thus exhibits a U-shaped pattern.

We further examine the shape of hazard rates in duration analyses. In a first step, we estimate the shape parameter ρ of a Weibull proportional hazards model. Depending on the parameter, the Weibull model can accommodate a decreasing, constant, or increasing baseline hazard. The estimate clearly points to a decreasing baseline hazard (see Table I in the Online Appendix). The data thus reject the case of a constant, time-invariant baseline hazard for the no-deadline treatment, predicted by our baseline model without memory limitations.

[16] Similar as in the main experiment, the within-deadline response rate is larger in D3 than in the D1 treatment. More generally, the within-deadline response rates in the follow-up experiment are non-monotonic in the deadline length. While this, again, conflicts with Proposition 1 from our baseline model, one has to keep in mind the limited power of the follow-up experiment (see fn. 11). The non-monotonicity nonetheless suggests that there might exist a 'sweet spot' of deadline length.

Figure 4: Daily Hazard Rates (Main Experiment)



Notes: The figure displays daily hazard rates for the treatments with a one-week deadline (top panel), three-week deadline (middle panel) and no deadline (bottom panel) over a period of six weeks (42 days) after sending the mailings.

Next, we turn to duration models that allow for time-varying treatment effects. Specifically, we estimate time-specific treatment effects by including separate treatment dummies for the first, second, and third week after the intervention, as well as a treatment dummy for all later weeks. The estimates document a strong impact of deadlines on the shape of hazard rates over time (see Online Appendix Table II). Consider first the D1 treatment. During the first week, the hazard rate is more than three times larger than in the ND treatment. However, as already observed in Figure 4, the effect of the short deadline is fully concentrated in the first week. For all later periods, the estimated hazard ratios are close to or smaller than one, illustrating the (partial) catch-up of response rates in the ND treatment relative to the D1 treatment in this phase (see Figure 1). For the three-week deadline, the estimates again show a strong impact of the deadline during the first week of the experiment. The estimates also confirm the U-shape in hazards observed in Figure 4: moving from the first to the second week, the hazard ratio shrinks substantially; from the second to the third week, hazards increase again. Both the initial decline as well as the later increase are statistically significant and qualitatively similar in specifications with additional controls (see Table II in the Online Appendix).

Hazard Rates and Distance to the Deadline. From a theoretical perspective, it is informative to compare treatment differences in hazards rates, holding constant the time that is left until the deadline. Remember that, according to the baseline model from Section 2, hazard rates in a given period should only depend on the time left until the deadline but not on how much time has passed since a patient received a postcard. Hence, without memory limitations, the hazard rate in the first week of the D1 treatment should be identical to the third-week hazard in the D3 treatment. In contrast, if memory limitations are relevant, we would expect week-3 hazards in the D3 treatment to be lower than week-1 hazards in the D1 treatment, as some patients may have turned inattentive in the meantime. The estimates reported in Table II clearly reject the case of identical pre-deadline hazards: in all specifications, the observed hazard in the pre-deadline week of the D1 treatment (i.e., during week 1) is significantly higher than the hazard during the pre-deadline week of the D3 treatment (i.e., week 3).

One might question the comparability of the pre-deadline weeks between the D1 and D3 treatment, since for the D1 treatment this period also coincides with the first week after the intervention. While this issue is irrelevant from a theoretical perspective, we can address this point by comparing differences in *daily* rather than weekly hazards. Doing so yields similar results as before. When we compare the daily pre-deadline hazards between treatments, we find that the empirically observed hazard rates are higher in the D1 than in the D3 treatment for each of the five days before the deadline (see Figure III in the Online Appendix).[17]

Follow-up Experiment. Our key findings on the shape of hazard rates in the main experiment are again corroborated when examining hazards in the follow-up experiment. First, for the D1, D3, and ND treatment, the hazard patterns in the follow-up experiment are strikingly similar to the ones observed in our main experiment: hazard rates are either decreasing (D1, ND, and EoY treatment) or have a U-shape (D3, D6, and D10; see Figure II in the Online Appendix). The U is considerably flattened out under the longer deadlines. In fact, for people facing a ten-week deadline (D10), there is no noticeable increase in hazard rates right before the deadline. This once more indicates that, under longer and longer deadlines, people's behavior approaches the limit case of the no-deadline treatment. Moreover, we again find that – similar to the main experiment – hazard rates in a given pre-deadline period depend negatively on the length of the deadline.[18]

Heterogeneity. The evidence reported above documents three key observations on the shape of hazard rates in our setting: (i) in treatments without a deadline, the baseline hazard decreases over time; (ii) hazard rates are

[17]To further back this observation, we estimate models with time-varying treatment effects on a daily rather than weekly basis. The results show that hazards for the last day (or the last two days) before the deadline are generally higher in the D1 than in the D3 treatment (see Table III in the Online Appendix).

[18]These results, as well as comprehensive duration analyses, are discussed in Section B.2 in the Online Appendix.

decreasing in the D1 treatment and exhibit a U-shaped pattern in the D3 treatment; (iii) in a given pre-deadline period, hazard rates are the lower, the more time has elapsed before. All these observations are consistent with the memory-limitations framework from Section 2 but at odds with the predictions from our baseline model. To judge whether one can indeed interpret these findings as further evidence supporting the relevance of memory limitations – or whether they are merely an artefact of differential responses of heterogeneous patients – we next examine the role of type heterogeneity in our setting.

As noted above, a possible consequence of type heterogeneity could be differential sorting of patients with different opportunity costs: types with lower opportunity costs may respond quickly, whereas high-cost types might respond later. This could, in principle, lead to decreasing or U-shaped hazard rates in the overall sample – even if hazards within each subgroup are monotonically increasing. In a first attempt to assess this concern, we follow Perry (1972) and study the shape of hazard rates among subgroups with similar socio-demographic characteristics. In addition, we exploit individual-level information that is usually unobserved (e.g., on patients' dental health or their frequency of past appointments). This exercise reveals no systematic differences in the shape of hazard rates across different subgroups (see Figures V–VII in the Online Appendix). In particular, none of the subgroups shows monotonically increasing pre-deadline hazards; instead, hazard rates are decreasing or U-shaped for all subgroups in all treatments.

Second, we use information on patients' responses to reminder postcards from the pre-intervention period to derive proxies for systematic differences in individuals' opportunity costs to respond. Our approach draws on data from an earlier study, which evaluated the basic impact of reminders at the same dental clinic (Altmann and Traxler, 2014).[19] Based on these data and a rich set of background characteristics, we first estimate which patient types have a particularly high propensity to quickly respond to a reminder. Using these estimates, we then predict the propensity that a patient-treatment observation in our main experiment yields a 'quick' response.[20] The predictions allow us to examine hazard rates separately for groups with similar immediate-response propensities – i.e., patient types with similar opportunity costs.

This exercise yields no indication of increasing hazards within patient groups with similar propensities to respond. When we compare hazard rates among the top quartile (i.e., patients with the highest predicted probability to respond quickly) and the remaining patients, we observe, consistently with our prediction approach,

[19] Altmann and Traxler (2014) randomized whether patients received a reminder and how the reminder was framed (without imposing any deadlines or economic incentives). In line with the relevance of memory limitations, they document strong reminder effects.

[20] More specifically, we fit a logit model using all available background variables (including some interacted variables), indicators for sending weeks within a month, and dummies for different sending quarters of the year on a sample of $N = 1,176$ patient-postcard observations. We define all responses within one week after a postcard was sent as 'quick'. To assess the robustness of our predictions, we also studied a LASSO logistic model (using the *glmnet* package). The shrunken model yields very similar predictions and split-sample results as those reported below. The same holds for using alternative estimation models (e.g., probit) or outcome variables (e.g., responses within two weeks).

that hazard rates are indeed higher for the high-propensity group (see Online Appendix Figure VIII); however, within both groups of patients – which should, by construction, be more homogeneous in terms of their opportunity costs – pre-deadline hazards are again decreasing (D1 and ND treatment) or U-shaped (D3).[21]

In a third step, we conduct a series of numerical simulations to assess the relevance of type heterogeneity for shaping population hazards in our setting (see Online Appendix C.3). Note that hazard rates on the first day of the intervention are not affected by differential sorting. We thus use day-1 hazards to calibrate (linear and degenerated) opportunity-cost distributions and hazard rates. Our simulations then consider different patient types that (a) match the empirically observed day-1 hazard rates of different subgroups and (b) imply non-decreasing hazard rates within each group (in line with our baseline model without memory limitations). Starting from these presumptions, we find no examples of empirically plausible type mixtures that give rise to decreasing or U-shaped population hazard rates. This is mainly due to the fact that the empirically observed day-1 hazards rates are, overall, rather low and do not differ dramatically across subgroups.[22] Both factors limit the influence of differential sorting. Counterfactual simulations show that it would require much stronger differences in initial hazards across subgroups and, additionally, an implausibly large group of highly responsive types such that differential sorting could result in U-shaped hazards under a 3-week deadline and decreasing hazards under a 1-week deadline. The most responsive subgroup, for instance, would need to display day-1 hazard rates that are at least twice as large as those documented in Figure VIII (see Online Appendix C.3).

To wrap up, our heterogeneity analyses do not detect a single subgroup with increasing pre-deadline hazards. Moreover, simulation exercises document that differential sorting should – given the observed levels and level differences in (early-period) hazards – only play a limited role in our context. While it is ultimately impossible to rule out that unobserved individual-level heterogeneity affects population hazards, our analysis suggests that heterogeneity is likely not the key driver of observations (i)–(iii) discussed above. When taken together with our findings on cumulative response rates, the hazard rate analysis suggests that memory limitations are indeed relevant in shaping individuals' reactions to deadlines.

5 Discussion

The results of our experiments document that deadlines constitute a powerful, low-cost tool to increase and accelerate task performance. A simple cost-effectiveness analysis – which is detailed in Section C.4 in the

[21] A similar picture emerges when we consider other subgroups of patients, e.g., the quartile with the lowest predicted propensity to quickly respond or the 'middle 50%'.

[22] For example, among the top-25% of patients with the highest predicted propensity to respond, the day-1 hazard rate under a three-week deadline (D3 treatment) is 0.08, whereas the corresponding number for the remaining patients is 0.04 (see Figure VIII in the Online Appendix).

Online Appendix – further illustrates this point. The analysis compares the different treatments according to the dentist's average costs to make a patient arrange a check-up appointment (i.e., the costs for printing, handling, and sending the reminder postcards and the input costs for the different rewards in terms of material and personnel; cp. Table VII in the Online Appendix). The simple back-of-the-envelope calculation provides several insights. First, it illustrates the value of deadlines, even if no explicit reward is attached to them. Considering all check-ups arranged in the no-reward conditions within 100 days, the higher response rates in the two deadline treatments imply a 10–15% lower costs per arranged check-up than the corresponding 'pure-reminder' treatment without deadlines and rewards. When considering response rates over shorter time periods (e.g., responses within one or three weeks), the comparison is even more favorable for the deadline treatments.

Another important observation is that, from a cost-effectiveness perspective, relatively short deadlines yield a further benefit: they lead patients to respond earlier, but do not lower the long-run response rates compared to the treatments with a longer or no deadline. Yet, the rewards only have to be paid for *timely* responses within the deadline – which is particularly relevant for the more expensive, large-reward conditions: with a deadline, the costs per arranged check-up are between 25% (D3) and 40% (D1) lower than in the corresponding ND treatment. Tying incentives to a reasonably short deadline thus has the potential to induce earlier and more responses at a lower overall cost than comparable open-ended incentives. However, this 'triple dividend' of deadlines only realizes with a sufficiently high post-deadline performance rate. Moreover, the gains from imposing deadlines in one domain (for one task) might lead to negative spillovers on other domains (other tasks) if cognitive resources are scarce (see, e.g., Mullainathan and Shafir, 2013; Altmann et al., 2021). Assessing the nature and potential consequences of these interdependencies is an important task for future research.

Our experiments also help to shed further light on the question of who does (and who does not) 'benefit' from facing a deadline. To examine this question, we compare cumulative response rates among different subgroups of our sample. The data indicate that imposing a deadline triggers a relatively *larger* increase in task performance for groups with *lower* response rates in the no-deadline environment. This point is illustrated in Figure IX in the Online Appendix, which compares the subgroups with the highest vs. lower predicted propensities of responding quickly. For the former group, the response rate after 100 days is 5.6% higher under the three-week deadline than in the no-deadline environment (with response rates of 48.9% vs. 46.3%, respectively). For the group with lower propensities to respond, the corresponding increase is 13.6% (response rates of 35.2% vs. 31.0%). In our context, deadlines thus seem particularly well suited for inducing task completion among subgroups with a lower baseline propensity to carry out the task at hand. This finding contributes to a growing body of research studying the incidence of behavioral interventions (see, e.g., Heffetz et al., 2021).

Complementing our evidence on the behavioral effects of deadlines, we also conducted a large online survey experiment on people's perceptions of deadlines (N = 3,078), as well as a smaller post-experimental survey at the dental clinic (N = 273). In the online survey, participants were randomly assigned to a vignette scenario in which they were shown one of the postcards from our experiment ('Imagine you receive the following postcard from your dentist...'). Participants who were confronted with different deadline lengths do not differ systematically in how they perceive (i) the dentist's competence, (ii) his economic or (iii) benevolent intentions ('wants me not to postpone/forget'), or (iv) the potential presence of an acute health problem (see Table X in Online Appendix C.5). In both surveys, however, between 80 and 95% of participants stated that the deadline on the postcard helps them not to postpone and forget arranging a check-up. We also asked survey participants which deadline length they would perceive as ideal in the check-up context. A large majority of participants in both surveys (74 and 90%, respectively) prefer a deadline of 4 weeks or less. These preferences match the 'success' of the (relatively) short deadline treatments in our experiments quite well, suggesting that individuals may – at least partially – anticipate their lower responsiveness to longer deadlines. At the same time, participants of the online survey who face a vignette scenario with a relatively tight deadline are also more likely to respond that the deadline puts them under pressure. Hence, while survey respondents generally appreciate tight deadlines, the latter do not come without costs. These costs ought to be carefully weighed against the potential benefits of deadlines in different applications.

6 Conclusions

This paper studied the interplay of deadlines and incentives in intertemporal choices. The results from a simple theoretical model, two natural field experiments, and complementary surveys suggest that individuals' responses to deadlines are shaped by limitations in memory and attention. We find that imposing deadlines leads to persistently higher task completion rates than having no deadline at all. Relatively tight deadlines can trigger earlier responses – without harming overall completion rates. From a managerial perspective, this also means that deadlines can constitute a powerful tool to increase the cost-effectiveness of performance-contingent rewards: tight eligibility rules may reduce costs and accelerate task completion, without leading to significant declines in overall performance. Moreover, the finding that deadlines 'work', even if no explicit reward is attached to them indicates that deadlines can induce implicit incentives for timely task completion, or help to plan and structure work tasks akin to planning prompts. From a public policy perspective, regulators need to be careful in assessing what constitute 'good' deadlines. If, for instance, a company offers seemingly generous extensions of consumer rights to its customers – e.g., very long deadlines for cancelations or product returns – this offer might ultimately not be consumer friendly, at least for individuals with cognitive limitations.

References

ALTMANN, S., A. GRUNEWALD, AND J. RADBRUCH (2021): "Interventions and Cognitive Spillovers," *Review of Economic Studies*, forthcoming.

ALTMANN, S. AND C. TRAXLER (2014): "Nudges at the Dentist," *European Economic Review*, 72, 19–38.

ARIELY, D. AND K. WERTENBROCH (2002): "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science*, 13, 219–224.

BALASUBRAMANIAN, N., J. LEE, AND J. SIVADASAN (2018): "Deadlines, Workflows, Task Sorting, and Work Quality," *Management Science*, 64, 1477–1973.

BERTRAND, M., D. KARLAN, S. MULLAINATHAN, E. SHAFIR, AND J. ZINMAN (2010): "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment," *Quarterly Journal of Economics*, 125, 263–306.

BESHEARS, J., K. L. MILKMAN, AND J. SCHWARTZSTEIN (2016): "Beyond Beta-Delta: The Emerging Economics of Personal Plans," *American Economic Review*, 106, 430–434.

BHARGAVA, S. AND D. MANOLI (2015): "Psychological Frictions and the Incomplete Take-up of Social Benefits: Evidence from an IRS Field Experiment," *American Economic Review*, 105, 3489–3529.

BISIN, A. AND K. HYNDMAN (2020): "Present-bias, Procrastination and Deadlines in a Field Experiment," *Games and Economic Behavior*, 119, 339–357.

BURGER, N., G. CHARNESS, AND J. LYNHAM (2011): "Field and online experiments on self-control," *Journal of Economic Behavior and Organization*, 77, 393 – 404.

CAMPBELL, A., F. EDERER, AND J. SPINNEWIJN (2014): "Delay and Deadlines: Freeriding and Information Revelation in Partnerships," *American Economic Journal: Microeconomics*, 6, 163–204.

CHETTY, R., J. N. FRIEDMAN, S. LETH-PETERSEN, T. H. NIELSEN, AND T. OLSEN (2014): "Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark," *Quarterly Journal of Economics*, 129, 1141–1219.

CHETTY, R., A. LOONEY, AND K. KROFT (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–1177.

DAMGAARD, M. T. AND C. GRAVERT (2017): "Now or Never! The Effect of Deadlines on Charitable Giving: Evidence from Two Natural Field Experiments," *Journal of Behavioral and Experimental Economics*, 66, 78–87.

DUSEK, L., N. PARDO, AND C. TRAXLER (2020): "Salience, Incentives, and Timely Compliance: Evidence from Speeding Tickets," Discussion Paper 2020/19, Max Planck Institute for Research on Collective Goods.

ENGLMAIER, F., A. SCHMÖLLER, AND T. STOWASSER (2018): "Price Discontinuities in an Online Market for Used Cars," *Management Science*, 64, 2473–2972.

ERICSON, K. M. (2011): "Forgetting we Forget: Overconfidence and Memory," *Journal of the European Economic Association*, 9, 43–60.

——— (2017): "On the Interaction of Memory and Procrastination: Implications for Reminders, Deadlines, and Empirical Estimation," *Journal of the European Economic Association*, 15, 692–719.

GABAIX, X. (2019): "Chapter 4 - Behavioral inattention," in *Handbook of Behavioral Economics - Foundations and Applications 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, North-Holland, vol. 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, 261 – 343.

GLIED, S. AND M. NEIDELL (2010): "The Economic Value of Teeth," *Journal of Human Resources*, 45, 468–496.

HAUSHOFER, J. (2015): "The Cost of Keeping Track," Working Paper, Princeton University.

HEFFETZ, O., T. O'DONOGHUE, AND H. S. SCHNEIDER (2021): "Reminders Work, But for Whom? Evidence from New York City Parking-Ticket Recipients," *American Economic Journal: Economic Policy*, forthcoming.

HEISS, F., D. MCFADDEN, J. WINTER, A. WUPPERMANN, AND B. ZHOU (2016): "Inattention and Switching Costs as Sources of Inertia in Medicare Part D," NBER Working Paper No. 22765.

HERWEG, F. AND D. MÜLLER (2011): "Performance of Procrastinators: On the Value of Deadlines," *Theory and Decision*, 70, 329–366.

HOLMAN, J. AND F. ZAIDI (2010): "The Economics of Prospective Memory," Working Paper, UC Berkeley.

HOSSAIN, T. AND J. MORGAN (2006): "...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay," *Advances in Economic Analysis & Policy*, 5.

KAHANA, M. J. (2012): *Foundations of Human Memory*, Oxford University Press, Oxford UK.

KARLAN, D., M. MCCONNELL, S. MULLAINATHAN, AND J. ZINMAN (2016): "Getting to the Top of Mind: How Reminders Increase Saving," *Management Science*, 62, 3393–3411.

KNOWLES, S. AND M. SERVATKA (2015): "Transaction Costs, the Opportunity Cost of Time and Procrastination in Charitable Giving," *Journal of Public Economics*, 125, 54–63.

LACETERA, N., D. G. POPE, AND J. R. SYDNOR (2012): "Heuristic Thinking and Limited Attention in the Car Market," *American Economic Review*, 102, 2206–2236.

LAIBSON, D. I. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–478.

LANG, W. P., M. M. FARGHALY, AND D. L. RONIS (1994): "The Relation of Preventive Dental Behaviors to Periodontal Health Status," *Journal of Clinical Periodontology*, 21, 194–198.

MULLAINATHAN, S. AND E. SHAFIR (2013): *Scarcity: Why Having Too Little Means So Much*, New York, NY: Time Books, Henry Holt & Company LLC.

O'DONOGHUE, T. AND M. RABIN (1999): "Incentives for Procrastinators," *Quarterly Journal of Economics*, 114, 769–816.

PERRY, G. (1972): "Unemployment Flows in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 3, 245–292.

ROGERS, T., K. L. MILKMAN, L. K. JOHN, AND M. I. NORTON (2015): "Beyond Good Intentions: Prompting People to Make Plans Improves Follow-through on Important Tasks," *Behavioral Science and Policy*, 1, 33–41.

ROTH, A. E. AND A. OCKENFELS (2002): "Last-minute Bidding and the Rules for Ending Second-price Auctions: Evidence from eBay and Amazon Auctions on the Internet," *American Economic Review*, 92, 1093–1103.

SAEZ-MARTI, M. AND A. SJÖGREN (2008): "Deadlines and Distractions," *Journal of Economic Theory*, 143, 153–176.

SALANT, S. W. (1977): "Search Theory and Duration Data: A Theory of Sorts," *Quarterly Journal of Economics*, 91, 39–57.

SHU, S. B. AND A. GNEEZY (2010): "Procrastination of Enjoyable Experiences," *Journal of Marketing Research*, 47, 933–944.

TAUBINSKY, D. (2014): "From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice," Working Paper, Harvard University.

TAUBINSKY, D. AND A. REES-JONES (2018): "Attention Variation and Welfare: Theory and Ecidence from a Tax Salience Experiment," *Review of Economic Studies*, 85, 2462–2496.

WEINSCHENK, P. (2016): "Procrastination in Teams and Contract Design," *Games and Economic Behavior*, 98, 264–283.

Appendix: Complementary Tables

Table A.1: Summary Statistics and Balancing Tests (Main Experiment)

	D1	D3	ND	F-Stats
Age	35.70	35.75	36.24	0.311
	(15.88)	(16.48)	(15.98)	[0.732]
Female	0.56	0.59	0.54	2.074
	(0.50)	(0.49)	(0.50)	[0.126]
Distance	12.75	10.10	11.63	1.572
	(43.94)	(34.27)	(42.06)	[0.208]
Housing price	8.02	8.06	8.01	0.454
	(1.14)	(1.13)	(1.10)	[0.635]
Private HI	0.19	0.19	0.21	0.606
	(0.39)	(0.39)	(0.41)	[0.546]
Family	0.17	0.15	0.16	0.462
	(0.38)	(0.36)	(0.37)	[0.630]
At risk	0.11	0.11	0.10	0.593
	(0.32)	(0.31)	(0.30)	[0.553]
Patient retention	4.51	4.11	4.27	4.886
	(2.74)	(2.65)	(2.58)	[0.008]
Past showup$^\flat$	0.73	0.70	0.74	1.455
	(0.45)	(0.46)	(0.44)	[0.228]
Pain$^\flat$	0.22	0.25	0.27	2.029
	(0.42)	(0.44)	(0.44)	[0.132]
N	907	879	875	

Notes: The table presents summary statistics (mean values, standard deviations in parentheses) for individuals' background characteristics, separately for each of the three deadline conditions. The overall sample size is N = 2,661. 'Distance' denotes the great-circle kilometer distance between a person's home address and the dentist. 'Housing price' is the average rental price at the participant's home address (euros per square meter, excluding utilities and dues). 'Family' is a dummy indicating that several household members took part in the study. 'Private HI' is a dummy indicating whether a patient is covered by private health insurance (0 for public health insurance). 'At risk' indicates that a participant is recommended a shortened check-up interval of 3 or 4 months. 'Patient retention' is the number of years since a participant first visited the dentist. The dummy 'Past showup' indicates whether a participant had at least one check-up in the year prior to a given treatment. 'Pain' indicates if a participant was exposed to a painful dental treatment in the past. The last two variables (labeled with $\flat$) are only available for 2,189 [Past showup] and 1,990 [Pain] observations, respectively. The final column reports F-statistics [and the corresponding p-values in brackets] from tests for treatment differences based on separate regressions of each of the characteristics on dummies for the different deadline conditions. For nine out of ten variables we detect no significant treatment differences.

Table A.2: Summary Statistics and Balancing Tests (Follow-up Experiment)

	D1	D3	D6	D10	ND	EoY	F-Stats
Age	38.65	36.56	35.96	34.94	37.08	38.55	1.127
	(15.92)	(17.12)	(16.59)	(14.92)	(16.21)	(17.81)	[0.345]
Female	0.53	0.58	0.58	0.48	0.62	0.48	1.949
	(0.50)	(0.49)	(0.49)	(0.50)	(0.49)	(0.50)	[0.084]
Distance	12.74	10.07	8.58	6.82	15.05	13.50	2.000
	(43.23)	(25.94)	(19.29)	(14.55)	(56.42)	(34.36)	[0.077]
Housing price	8.09	8.12	8.13	8.14	7.99	7.86	1.029
	(1.18)	(1.07)	(1.21)	(1.07)	(1.23)	(1.14)	[0.400]
Private HI	0.18	0.20	0.23	0.17	0.17	0.25	0.774
	(0.38)	(0.40)	(0.42)	(0.38)	(0.38)	(0.44)	[0.569]
Family	0.12	0.17	0.19	0.16	0.15	0.14	0.500
	(0.32)	(0.38)	(0.39)	(0.36)	(0.36)	(0.35)	[0.776]
At risk	0.12	0.15	0.06	0.15	0.10	0.11	1.688
	(0.32)	(0.36)	(0.24)	(0.36)	(0.30)	(0.31)	[0.135]
Patient retention	4.63	4.66	5.12	5.07	4.78	4.80	0.699
	(3.02)	(2.84)	(2.78)	(2.93)	(2.96)	(3.32)	[0.624]
Past showup$^\flat$	0.73	0.73	0.77	0.78	0.74	0.68	0.524
	(0.44)	(0.44)	(0.42)	(0.42)	(0.44)	(0.47)	[0.758]
Pain$^\flat$	0.24	0.20	0.20	0.27	0.24	0.27	0.576
	(0.43)	(0.40)	(0.40)	(0.44)	(0.43)	(0.45)	[0.719]
N	169	163	163	166	175	91	

Notes: The table presents summary statistics (mean and standard deviation in parenthesis) of participant characteristics for each treatment of the follow-up experiment. The last column reports F-statistic [and the corresponding p-values in brackets] from tests for treatment differences based on separate regressions of each of the characteristics on a full set of treatment dummies. Sample size is $N = 927$. The two variables indicated with $\flat$, *Pain* (records on painful dental treatments in the past) and *Past-Showup* (one show-up in the year before the intervention) are only available for 716 and 789 observations, respectively.

Online Appendix

DEADLINES AND MEMORY LIMITATIONS

S. Altmann, C. Traxler, P. Weinschenk

A Theoretical Analysis

This section derives Predictions 1 and 2 from the main text. Extensions of the model are discussed in Section B below. We start by formally specifying the model sketched in the main text. The costs c_t of performing the task are drawn at the beginning of each period $t \in \{1, \ldots, T\}$ from a continuous distribution F with support $[0, \bar{k}]$. After observing the realization of costs in period t, the agent decides whether to carry out the task in this period or not. If the task is carried out, she obtains a reward with present value y, where $0 < y < \bar{k}$, such that her payoff is $y - c_t$.[1] If the task is not carried out and the deadline has not yet been reached, the agent advances to the next period $t + 1$. The decision problem ends once the task is performed or the deadline T is reached. If not explicitly stated differently, T is assumed to be finite. The discount factor is denoted by δ and between zero and one.

A.1 Case without Memory Limitations

Key Properties of the Option Value

We first examine how the agent's option value evolves over time. We claim that the option value is decreasing over time for any finite deadline. Formally, $V_t > V_{t+1}$ holds for all periods $t \in \{1, \ldots, T\}$. This property is intuitive: when there is more time left until the deadline, the agent has more chances to complete the task, which positively affects the agent's expected payoff, i.e., the option value.

PROOF: Combining (1) and (4) yields

$$V_t = \int_0^{\hat{c}_t} (\hat{c}_T - c) dF(c) + \delta V_{t+1}. \tag{A.1}$$

Since the decision problem ends at the latest after period T, the option value of reaching a period after T is zero: $V_t = 0$ for all $t > T$. Hence, the cost threshold in period T is $\hat{c}_T = y > 0$; see equation (1). From (A.1) we thus see that $V_T > 0$. The option value thus decreases from period T to period $T + 1$.

Next, consider the remaining periods. Because the agent chooses the cost threshold in every period t to maximize V_t, the Envelope Theorem together with (4) implies that the agent's current option value V_t benefits from a higher future option value V_{t+1}:

$$\frac{dV_t}{dV_{t+1}} = \frac{\partial V_t}{\partial \hat{c}_t} \cdot \frac{\partial \hat{c}_t}{\partial V_{t+1}} + \frac{\partial V_t}{\partial V_{t+1}} = \frac{\partial V_t}{\partial V_{t+1}} = (1 - F(\hat{c}_t))\delta > 0, \tag{A.2}$$

[1] If the reward Y is received $\ell > 0$ periods after the task is performed, $y = \delta^\ell Y$. In the case of a one-period delay, we have $y = \delta Y$ and the cost threshold in equation (1) would be $\hat{c}_t = \delta(Y - V_{t+1})$.

where the inequality follows since $\hat{c}_t \leq y$. From equations (1) and (A.1) we see that if in period $T-1$ the future option value would be the same as in period T, i.e., $V_T = V_{T+1}$, the current option values would be the same, too, $V_{T-1} = V_T$. But from above we know that $V_T > V_{T+1}$. Together with (A.2), this implies that $V_{T-1} > V_T$. Repeating these arguments for earlier periods yields that

$$V_t > V_{t+1} \text{ for all } t \in \{1, \ldots, T\}. \tag{A.3}$$

□

We next establish that the option value is always such that probability that the agent completes the task in a certain period, given that she has not completed the task yet, is strictly between zero and one. Formally, in all periods $t \leq T$, under the optimal decision rule, the hazard rate is

$$h_t = F(\hat{c}_t) \in (0, 1). \tag{A.4}$$

PROOF: Note first that if $F(\hat{c}_t) = 0$, then $\hat{c}_t = 0$. Due to (4), we then have $V_t = \delta V_{t+1} \leq V_{t+1}$, which contradicts inequality (A.3). If $F(\hat{c}_t) = 1$, then $\hat{c}_t > y$. By (1), however, this requires that $V_{t+1} < 0$, which cannot hold since inequality (A.3) and $V_{T+1} = 0$ imply that all option values are non-negative. In case the deadline is infinite, $T \to \infty$, the option value is $V_t = \bar{V} \in (0, y)$, see below, which implies that $F(\hat{c}_t) \in (0, 1)$ holds also in this case. □

We finally establish that there exists an upper bound on the option values $\bar{V}$ and that this upper bound is reached if the deadline is infinitely long. Formally, $V_t < \bar{V}$ for all deadlines $T < \infty$ and $\lim_{T \to \infty} V_t = \bar{V}$.

PROOF: Consider some period $t \leq T$. By (A.1),

$$V_t - V_{t+1} = \int_0^{\hat{c}_t} (\hat{c}_T - c) dF(c) - (1 - \delta) V_{t+1}. \tag{A.5}$$

Using (1) and (A.5), let $\bar{V}$ solve

$$0 = \int_0^{y - \delta \bar{V}} (y - \delta \bar{V} - c) dF(c) - (1 - \delta) \bar{V}. \tag{A.6}$$

By the Intermediate Value Theorem and monotonicity, a unique solution $\bar{V}$ exists and $\bar{V} \in (0, y)$. Recall that $V_{T+1} = 0$ and note that by (A.2)

$$\frac{d(V_t - V_{t+1})}{dV_{t+1}} = (1 - F(\hat{c}_t))\delta - 1, \tag{A.7}$$

which is strictly between -1 and $\delta - 1$. Therefore, if we start at period $T + 1$ and then go back in time, the option value V_t increases, approaches $\bar{V}$, but never exceeds $\bar{V}$. The option value V_t is thus bounded and approaches the bound $\bar{V}$ as the deadline goes to infinity:

$$V_t < \bar{V} \text{ for all } T < \infty \text{ and } \lim_{T \to \infty} V_t = \bar{V}. \tag{A.8}$$

□

Hazard Rate – Prediction 1 (iii)

We first explore how the hazard rate evolves over time. Since the option value V_t is decreasing over time (see inequality A.3) and the cost-threshold $\hat{c}_t$ is decreasing in the option value (see equation 1), the cost-threshold is increasing over time. It follows from (2) that the hazard rate is hence increasing over time as well:

$$h_t < h_{t+1} \ \text{ for all } \ t \in \{1, \dots, T-1\}. \tag{A.9}$$

Intuitively, the closer the deadline, the higher are the costs that the agent is willing to bear to get the task done. This in turn translates into a higher hazard rate. Note that this result contrasts with the case where the agent faces no deadline/an infinitely long deadline. With an infinitely long deadline, $T \to \infty$, the environment is time-invariant and the hazard rate is constant over time: $h_t = F\left(y - \delta \bar{V}\right)$.

We next explore how the option value and the hazard rate depend on how much time is left until the deadline. Since the option value of reaching the period after the deadline is zero (i.e., $V_{T+1} = 0$), (1) and (4) imply that the option value V_T is the same for all possible deadlines T. Repeating this argument shows that the option values in preceding periods, $V_{T-1}, V_{T-2}, \dots$, are the same for all T, too. Hence, if we vary the length of the deadline, the option value only depends on the time left until the deadline but *not* on the absolute value of time (i.e., the time that has passed since period 1):

$$V_t|_{T=T'} = V_{t+s}|_{T=T'+s} \ \text{ for all } \ T' \in \mathbb{N} \ \text{ and } \ s \in \mathbb{N}. \tag{A.10}$$

Using (1), we can make the same statement for the cost threshold and, hence, for the hazard rate:

$$h_t|_{T=T'} = h_{t+s}|_{T=T'+s} \ \text{ for all } \ T' \in \mathbb{N} \ \text{ and } \ s \in \mathbb{N}. \tag{A.11}$$

Within-Deadline Performance Rate – Prediction 1 (i)

We now analyze how the within-deadline performance rate is related to the length of the deadline. Using (A.11) and (A.4) we obtain that

$$p(T') = 1 - \Pi_{t=1}^{T'}(1 - h_t|_{T=T'}) = 1 - \Pi_{t=T''-T'+1}^{T''}(1 - h_t|_{T=T''}) < 1 - \Pi_{t=1}^{T''}(1 - h_t|_{T=T''}) = p(T'') \tag{A.12}$$

holds for all pairs of deadlines T' and T'', where $T'' > T'$ and T'' is finite. The probability that the agent performs the task within the deadline is therefore higher under the long deadline T'' than under the short deadline T'. Intuitively, during the last T' periods before the deadline, the agent's optimal strategy is the same under the short and the long deadline. For the long deadline T'', however, the agent may have performed the task already in periods $t \leq T'' - T'$.

The result also holds extends to the case with an infinitely long deadline: $p(T') < p(T'')$ for all pairs of deadlines T' and T'', where T' is finite and T'' is infinite, since $p(T') < 1$ and $p(T'') = 1$ by (3) and (A.4).

We can thus conclude that the within-deadline performance rate is increasing in the length of the deadline:

$$p(T') < p(T'') \tag{A.13}$$

holds for all pairs of deadlines T' and T'', where $T'' > T'$ and T'' could be finite or infinite.

Cumulative Performance Rate – Prediction 1 (ii)

We next explore how the cumulative performance rates depend on the deadlines. Consider an arbitrary pair of deadlines T' and T'', where $T'' > T'$ and T'' could be finite or infinite. The hazard rate is higher with the short deadline T' than with the long deadline T'' in all periods $t \leq T'$ by (A.9) and (A.11) in case T'' is finite and since $V_t < \bar{V}$ for all deadlines $T < \infty$ and $\lim_{T\to\infty} V_t = \bar{V}$ by (1) and (2) in case T'' is infinite. Equation (3) then directly implies that

$$p(t)|_{T=T'} > p(t)|_{T=T''} \quad \text{for all} \quad t \leq T'. \tag{A.14}$$

The intuition why the cumulative performance rates are higher under a short deadline than under a long deadline in early periods, is that there is more time pressure with the short deadline, such that the agent has a higher willingness to carry out the task – i.e., she uses higher cost thresholds – which makes task performance more likely.

We next show that long deadlines overturn short deadlines in terms of cumulative performance in later periods. Formally, for periods $t \geq T''$, the cumulative performance rates equal the within-deadline performance rates, i.e., $p(t)|_{T=T'} = p(T')$ and $p(t)|_{T=T''} = p(T'')$, such that by (A.13)

$$p(t)|_{T=T'} < p(t)|_{T=T''} \quad \text{for all} \quad t \geq T''. \tag{A.15}$$

Since in the intermediate periods $t \in (T', T'')$, the cumulative performance rate increases over time for the long deadline T'', while it stays constant for the short deadline T' at $p(T')$, there exists a date $\tau \in (T', T'')$ such that

$$p(t)|_{T=T'} > p(t)|_{T=T''} \quad \text{for all} \quad t < \tau \quad \text{and} \quad p(t)|_{T=T'} < p(t)|_{T=T''} \quad \text{for all} \quad t > \tau. \tag{A.16}$$

Thus, the cumulative performance rates $p(t)$ are higher in early periods $t < \tau$ with the short deadline T' than with the long deadline T'', while the result reverses in later periods $t > \tau$.

A.2 Case with Memory Limitations

We now consider an agent with memory limitations.[2] At the end of each period t, the agent turns inattentive with probability $\gamma \in (0, 1]$.[3] For the moment, we further assume that the agent is naive about her memory limitations. The case of a sophisticated agent is discussed in Section B.2 below. Being naive, the agent does not foresee the possibility that she might become inattentive. She therefore uses the same option values and cost thresholds as in the case without memory limitations. From the perspective of period 1, the probability

[2]When we refer to 'memory limitations', this subsumes the case of simple forgetting as well as more complex processes of being inattentive to a task that would require one's effort or attention to be carried out. For excellent overviews of research on memory and (in)attention, see Kahana (2012) and Gabaix (2019).

[3]A similar assumption on memory limitations is made in Ericson (2017). Allowing for time-dependent inattention probabilities γ_t or the possibility of 'recalling' the task again at some later point in time complicates notation, without changing the key qualitative insights. Attention dynamics in settings involving repeated tasks are further analyzed in Taubinsky (2014).

that the agent performs the task in period t is then given by

$$q_t = F(\hat{c}_t)(1-\gamma)^{t-1} \prod_{s=1}^{t-1} (1 - F(\hat{c}_s)). \tag{A.17}$$

The probability that the agent performs the task in period t thus depends on (i) the probability of the cost draw in this period being low enough, (ii) the probability of the agent still being attentive in period t, and (iii) the probability of not having performed the task in some earlier period $s < t$. The hazard rate for an agent with limited memory is then given by

$$h_t = \frac{q_t}{1 - \sum_{s=1}^{t-1} q_s}. \tag{A.18}$$

Hazard Rate – Prediction 2 (iii)

Hazard Rate over Time. To determine how the hazard rate evolves over time, we consider the hazard ratio

$$\frac{h_{t+1}}{h_t} = \frac{\frac{q_{t+1}}{1-\sum_{s=1}^{t} q_s}}{\frac{q_t}{1-\sum_{s=1}^{t-1} q_s}}, \tag{A.19}$$

which can be rewritten as

$$\frac{h_{t+1}}{h_t} = \underbrace{\frac{F(\hat{c}_{t+1})}{F(\hat{c}_t)}}_{\text{(A) Option-value effect}} \underbrace{(1-\gamma)(1-F(\hat{c}_t)) \frac{1 - \sum_{s=1}^{t-1} F(\hat{c}_s)(1-\gamma)^{s-1} \prod_{r=1}^{s-1}(1-F(\hat{c}_r))}{1 - \sum_{s=1}^{t} F(\hat{c}_s)(1-\gamma)^{s-1} \prod_{r=1}^{s-1}(1-F(\hat{c}_r))}}_{\text{(B) Memory-limitation effect}}. \tag{A.20}$$

This expression illustrates two countervailing effects that shape the behavior of a memory limited agent. The first fraction on the right-hand side captures the *option-value effect*. Since the option value is decreasing over time (see inequality A.3), the cost threshold is higher in $t+1$ than in t and hence $F(\hat{c}_{t+1}) > F(\hat{c}_t)$. This effect contributes towards increasing the hazard over time. The remaining terms on the right-hand side – which are all positive and below one – capture the *memory-limitation effect*, which works into the opposite direction: due to the agent's memory limitations, the likelihood of having the task on the top of her mind is decreasing from one period to the next. The interplay of this effect with the former determines how the hazard evolves over time.

We next investigate when which effect dominates. Recognize that by (1) and (A.3) the fraction that captures the option-value effect (Term A) is strictly greater than one and finite. Since the terms that capture the memory-limitation effect are decreasing in the inattention probability γ, approach one for $\gamma \to 0$ and zero for $\gamma \to 1$, by the Intermediate Value Theorem, there exists a unique threshold $\hat{\gamma}_t \in (0,1)$ such that

$$\begin{aligned} h_{t+1} &< h_t \text{ if } \gamma > \hat{\gamma}_t, \\ h_{t+1} &= h_t \text{ if } \gamma = \hat{\gamma}_t, \\ h_{t+1} &> h_t \text{ if } \gamma < \hat{\gamma}_t. \end{aligned} \tag{A.21}$$

Hence, when γ is sufficiently high, the memory-limitation effect dominates and the hazard rate is locally decreasing. Vice versa, if γ is sufficiently low, the option-value effect dominates and the hazard rate is locally increasing.

We now show that also the time until the deadline is a crucial factor that determines the shape of the hazard rates. From (A.7) and the subsequent arguments it follows that the difference in option values V_t and V_{t+1} is larger, the lower is V_{t+1}, i.e., the closer period t is to the deadline T. This implies that the option-value effect is relatively strong in periods close to the deadline, whereas it is weaker in early periods where the deadline is still far away. The option-value effect is thus more likely to be dominated by the memory-limitation effect in periods long before the deadline. During early periods, hazard rates are thus more likely to be locally decreasing. Vice versa, the option-value effect is more likely to dominate in periods close to the deadline. As a result, agents with memory limitations may exhibit U-shaped hazard functions that are initially decreasing but then increase in later periods once the deadline comes closer.

Finally, for the case without a deadline, $T \to \infty$, the option value is constant over time and, consequently, the option-value effect plays no role. As a result, the memory-limitation effect always dominates and the hazard rate is monotonically decreasing over time.

Hazard Rate and the Absolute Value of Time. In contrast to the case without memory limitations, the hazard rate for an agent with limited memory depends negatively on the absolute value of time (i.e., the time that has passed since $t = 1$), holding the distance to the deadline fixed. To see this, we compare the hazard rate of period t in case of some deadline T' to that of period $t+1$ in case of deadline $T'+1$. From equation (A.18) we obtain

$$h_t|_{T=T'} = \frac{F(\hat{c}_t|_{T=T'})(1-\gamma)^{t-1}\prod_{s=1}^{t-1}(1-F(\hat{c}_s|_{T=T'}))}{1-\sum_{s=1}^{t-1}F(\hat{c}_s|_{T=T'})(1-\gamma)^{s-1}\prod_{r=1}^{s-1}(1-F(\hat{c}_r|_{T=T'}))}, \tag{A.22}$$

$$h_{t+1}|_{T=T'+1} = \frac{F(\hat{c}_{t+1}|_{T=T'+1})(1-\gamma)^{t}\prod_{s=1}^{t}(1-F(\hat{c}_s|_{T=T'+1}))}{1-\sum_{s=1}^{t}F(\hat{c}_s|_{T=T'+1})(1-\gamma)^{s-1}\prod_{r=1}^{s-1}(1-F(\hat{c}_r|_{T=T'+1}))}. \tag{A.23}$$

Applying (1) and (A.10) yields

$$h_t|_{T=T'} > h_{t+1}|_{T=T'+1} \quad \text{for all } \; T' \in \mathbb{N}. \tag{A.24}$$

Holding the distance to the deadline fixed, the absolute value of time thus negatively affects the hazard rate. This is intuitive: with more time, it is more likely that the agent has turned inattentive and hence less likely that she completes the task.

Within-Deadline Performance Rate – Prediction 2 (i)

A further interesting deviation from our benchmark results is obtained when we examine the within-deadline performance rates. Consider first the case where the inattention probability γ is small, i.e., $\gamma \to 0$. In this case, the hazard rates with memory limitations approach those without memory limitations. As a consequence, cf. (A.13), the within-deadline performance rate $p(T)$ is then increasing in the length of the deadline.

The result looks quite different, however, when γ is large, $\gamma \to 1$. In this case, the hazard rate is positive in the first period and approaches zero in all subsequent periods, cf. (A.17) and (A.18). The within-deadline performance rate $p(T)$ thus approaches the first-period hazard rate $F(\hat{c}_1)$. As the cost threshold $\hat{c}_1$ is decreasing in the length of the deadline T,[4] the within-deadline performance rate $p(T)$ is decreasing in T, too. A shorter deadlines thus causes a higher within-deadline performance rate.

We can be more general. First note that the within-deadline performance rate can be expressed as $p(T) = \sum_{s=1}^{T} q_s$. From (A.17) we see that the probability that the agent performs the task in period s, measured from the perspective of $t = 1$, q_s is monotonically decreasing in the inattentive rate γ. Thus, the within-deadline performance rate $p(T)$ is monotonically decreasing in γ. Together with the above findings for $\gamma \to 0$ and $\gamma \to 1$ and the Intermediate Value Theorem, we get the following result: For any pair of deadlines T' and T'', where $T'' > T'$ and T'' could be finite or infinite, there exists a unique threshold $\dot{\gamma} \in (0, 1)$ such that

$$\begin{aligned} p(T'') &< p(T') \text{ if } \gamma > \dot{\gamma}, \\ p(T'') &= p(T') \text{ if } \gamma = \dot{\gamma}, \\ p(T'') &> p(T') \text{ if } \gamma < \dot{\gamma}. \end{aligned} \tag{A.25}$$

Thus, if the level of memory limitation is rather high, performance within the deadline is more likely for short rather than long deadlines, while it is vice versa for low levels of memory limitations.

To illustrate this result, let $c \sim U[0, 1]$, $y \in (0, 1)$, $\delta = 1$, $T' = 1$, and $T'' = 2$. Then $p(T') = y$ and $p(T'') = 1 - (1 - y + y^2/2)(1 - (1 - \gamma)y)$. The within-deadline performance rate with the short deadline T' is thus higher [equal, lower] than with the long deadline T'' if $\gamma > [=, <]\dot{\gamma} = \frac{y^2-3y+2}{y^2-2y+2}$.

Cumulative Performance Rate – Prediction 2 (ii)

As in the framework without memory limitation, the cumulative performance rate $p(t)$ is higher under a short deadline T' than under a long deadline T'' for any period $t \leq T'$, where $T'' > T'$ and T'' could be finite or infinite. This holds since in all these periods the cost thresholds and thus the hazard rates are higher under the short than under the long deadline due to (A.9) and (A.11) in case T'' is finite and due to (1) and (A.8) in case T'' is infinite.

In contrast to the case without memory limitations, however, it may no longer be true that the cumulative performance rate under a long deadline catches up in later periods and overtakes the one under a short deadline after some threshold period τ (where $T' < \tau < T''$). In particular, using that the cumulative performance rate is $p(t) = \sum_{s=1}^{t} q_s$ and that the within-deadline performance rate is $p(T) = \sum_{s=1}^{T} q_s$, we directly obtain by help of (A.17) that all cumulative performance rates are higher with the short deadline than with the long

[4]This follows directly from the facts that the option value V_2 is increasing in the length of the deadline T, cf. (A.3), and that the cost threshold $\hat{c}_1$ is decreasing in the option value V_2, cf. (1).

deadline whenever the within-deadline performance rate is higher with the short deadline than with the long deadline. Formally, if $p(T') > p(T'')$, then

$$p(t)|_{T=T'} > p(t)|_{T=T''} \tag{A.26}$$

holds for any period t. Recalling that $p(T') > p(T'')$ holds whenever $\gamma > \dot{\gamma}$ by (A.25), we thus obtain the following result: With sufficiently strong memory limitations, $\gamma > \dot{\gamma}$, the cumulative performance rates are persistently higher with a short deadline than under a long (possibly infinite) deadline.

B Extensions of the Model Framework

We next show that the simple model framework developed in Section 2 can be extended in a number of ways. First, we investigate the possibility of post deadline performance. We formalize this by letting the agent receive a high reward if she completes the task within the deadline and a relatively lower, but positive reward if she completes the task afterwards. There are thus still incentives to complete the task within the deadline. We show that due to the positive reward differential, the agent is in case of no memory limitations more likely to perform in any period before the deadline than in any period after the deadline, given that she has not yet performed the task before. Moreover, the hazard rate is increasing towards the deadline, since there is less and less time left for the agent to obtain the relatively high reward. We furthermore show that also the other results we obtained in the simple model still hold in the environment with the possibility of post deadline performance.

Second, we examine the case where the agent has memory limitations and is fully sophisticated, i.e., perfectly aware of her limitations. We show that a sophisticated agent adjusts her decision rules towards the probability of becoming inattentive. In particular, a sophisticated agent is, all else equal, willing to bear higher costs of task performance than a naive agent with memory limitations or an agent without memory limitations. The main effects caused by memory limitations are shown to be the same with sophistication as with naïveté.

Thereafter, we explore present-biased preferences and overoptimism. We then discuss the effects of deadlines on agent's well-being and, finally, examine the case of recall as well as additional costs for attending a check-up appointment.

B.1 Task Performance after the Deadline

Our framework naturally extends to environments where the agent can still perform the task after the deadline.[5] Analyzing post-deadline rewards is important in our setting, since the participants in the experiment can (and do) call the clinic after the deadline has passed. To formalize this, suppose that the agent receives the reward $\bar{y}$ if she completes the task within the deadline and $\underline{y}$ if she does so afterwards, where $\bar{y} \geq \underline{y} > 0$. In what follows,

[5]Note that environments where the reward is the same for performance before or after the 'deadline' are isomorphic to the case without a deadline, $T \to \infty$.

we show that the previously obtained results are also valid in this extended environment. Since post deadline performance is only relevant in case the deadline is finite, we concentrate on finite deadlines henceforth.

Case Without Memory Limitations. We are interested in the structure of the cost thresholds the agent uses, which in turn depend on the structure of the option values. Suppose initially that $\bar{y} = \underline{y}$, such that the agent uses a constant cost threshold for all periods, resulting in hazard rates that are also constant over time. Suppose now that $\bar{y}$ increases, whereas $\underline{y}$ remains constant. This leaves the cost thresholds and, consequently, the hazard rates after the deadline T constant. But how do the cost thresholds and hazard rates before the deadline change? From equation (4) and the Envelope Theorem it follows that

$$\frac{dV_T}{d\bar{y}} = F(\hat{c}_T) \in (0,1). \tag{A.27}$$

Using this insight and again equation (4) and the Envelope Theorem, we get that

$$\frac{dV_{T-1}}{d\bar{y}} = F(\hat{c}_{T-1}) + (1 - F(\hat{c}_{T-1}))\delta\frac{dV_T}{d\bar{y}} = F(\hat{c}_{T-1}) + (1 - F(\hat{c}_{T-1}))\delta F(\hat{c}_T), \tag{A.28}$$

which is again strictly between zero and one. Repeating the arguments leads to the insight that

$$\frac{dV_t}{d\bar{y}} \in (0,1) \text{ for all } t \in \{1, \ldots, T\} \tag{A.29}$$

and thus by (1) and (2) that for all pairs $\bar{y}$ and $\underline{y}$ with $\bar{y} > \underline{y}$

$$h_t > h_s \text{ for all } t \in \{1, \ldots, T\} \text{ and } s \geq T+1. \tag{A.30}$$

This is intuitive: if the agent is incentivized to perform the task before the deadline through a positive reward differential $\bar{y} - \underline{y} > 0$, she is more likely to complete the task in any period before the deadline than in any period after the deadline, given that she has not yet performed the task.

By essentially the same arguments as in the basic model, it also follows that the hazard rate is increasing over time until period T,

$$h_t < h_{t+1} \text{ for all } t \in \{1, \ldots, T-1\}, \tag{A.31}$$

and that the hazard rates depend only on the time left until the deadline, but not on the absolute value of time,

$$h_t|_{T=T'} = h_{t+s}|_{T=T'+s} \text{ for all } T' \in \mathbb{N} \text{ and } s \in \mathbb{N}. \tag{A.32}$$

We can thus conclude that Prediction 1(iii) still holds when task performance is still possible after the deadline.

It remains to show that also parts (i) and (ii) of Prediction 1 hold in this environment. Since hazard rates are positive and only depend on the time left until the deadline, but not on the absolute value of time, it directly follows that the within-deadline performance rate is still increasing in the length of the deadline. In a final step, we show that our results for the cumulative performance rates (summarized in equation A.16) also hold when task performance after the deadline is possible. Consider some arbitrary pair of deadlines T' and T'', with $T' < T''$. First, by (A.31) and (A.32), the hazard rates in all periods $t \leq T'$ are lower with the long deadline T'' than with the short deadline T', such that by (3) the cumulative performance rates in all periods $t \leq T'$ satisfy $p(t)|_{T=T'} > p(t)_{T=T''}$. Second, since the hazard rates in the last T' periods before the deadline are the same for both deadlines T' and T'' (see condition A.32), and the hazard rates after the deadline T' are lower than the hazard rates in the first $T'' - T'$ periods in case of deadline T'' (see condition A.30), equation (3) implies that the cumulative performance rate in period T'' is higher for the long than for the short deadline: $p(T'')|_{T=T'} < p(T'')_{T=T''}$. Finally, since the hazard rates under the long deadline T'' are higher for all periods between T' and T'' than those under the short deadline T' (see condition A.30), there must exist a threshold period $\bar{\tau} \in (T', T'')$ such that $p(t)|_{T=T'} > p(t)_{T=T''}$ for all $t < \bar{\tau}$ and $p(t)|_{T=T'} < p(t)_{T=T''}$ for all $t > \bar{\tau}$.

Case With Memory Limitations. The effects of memory limitations is not substantially affected by the possibility and the incentives to perform the task after the deadline. In particular, the formal analysis stays basically unchanged. As a consequence, Prediction 2 stays valid.

B.2 Sophistication about Memory Limitations

So far, we have assumed that an agent with memory limitations is naive, in the sense of not aware that memory limitations may cause her to become inattentive (compare Holman and Zaidi, 2010). We next examine the case where the agent has memory limitations and is fully sophisticated, i.e., is perfectly aware of the limitations. Note that the arguments are similar for the case of partial sophistication; the agent's perceived probability of becoming inattentive, $\hat{\gamma}$, is then positive, but lower than the true rate of inattention, $0 < \hat{\gamma} < \gamma$.

If the agent does not complete the task in period t, the present value of her payoff is $(1-\gamma)\delta V_{t+1}^{S}$, where we use the superscript S to denote the case of sophistication. Hence, her optimal decision rule is to complete the task in period t if and only if

$$c_t \leq \hat{c}_t^S := y - (1-\gamma)\delta V_{t+1}^S. \tag{A.33}$$

In contrast to a naive agent, the sophisticated agent thus adjusts her decision rule to the inattention probability γ. In particular, when the agent decides on whether she performs the task in period t, she takes into account that it might otherwise drop off her mind. Her cost threshold for performing the task in t is thus increasing in γ. All else equal, the cost threshold of a sophisticated agent is therefore higher than for a naive agent or an agent without memory limitations, who both decide based on $\gamma = 0$.

Intuitively, one might expect that sophistication completely revokes the effects of memory limitations, such that the hazard rate h_t^S is again increasing for all $t \in \{1, \ldots, T\}$. This conjecture, however, turns out to be wrong. Let us illustrate this point for the case of $T = 2$. In the second period, the agent's cost threshold for performing the task is $\hat{c}_2 = y$. This holds for both the naive and sophisticated agent. The second-period hazard rate is then $h_2 = h_2^S = (1-\gamma)F(y)$, which is again independent of the agent's naïveté or sophistication. From above we know that in period $t = 1$, the sophisticated agent anticipates the possibility that she might become inattentive at the end of the period. She therefore applies a higher cost threshold in $t = 1$ than a naive agent. Consequently, the first-period hazard rate is higher for a sophisticated than the one for a naive agent.[6] Hence, whenever the hazard rate is decreasing over time for a naive agent, the same must also hold true for a sophisticated agent. Formally, if $h_1 > h_2$ then $h_1^S > h_2^S$. Below we explore these points analytically.

Formal Analysis of Sophisticated Agents. The hazard rate in period t is

$$h_t^S = \frac{F(\hat{c}_t^S)(1-\gamma)^{t-1}\prod_{s=1}^{t-1}(1-F(\hat{c}_s^S))}{1-\sum_{s=1}^{t-1}F(\hat{c}_s^S)(1-\gamma)^{s-1}\prod_{r=1}^{s-1}(1-F(\hat{c}_r^S))}. \tag{A.34}$$

Under the decision rule (A.33), the option value – given that the task is still on the top of the agent's mind – is

$$V_t^S = \int_0^{\hat{c}_t^S}(y-c)dF(c) + \left(1-F\left(\hat{c}_t^S\right)\right)(1-\gamma)\delta V_{t+1}^S. \tag{A.35}$$

By the Envelope Theorem, the sophisticated agent benefits from a higher option value, $dV_t^S/dV_{t+1}^S > 0$. Together with the same line of arguments as in case of no memory limitations, this implies that a sophisticated agent benefits from a longer deadline.

To explore how the hazard rate evolves over time, we need to adjust the hazard ratio from equation (A.19):

$$\frac{h_{t+1}^S}{h_t^S} = \frac{F(\hat{c}_{t+1}^S)}{F(\hat{c}_t^S)}(1-\gamma)(1-F(\hat{c}_t^S))\frac{1-\sum_{s=1}^{t-1}F(\hat{c}_s^S)(1-\gamma)^{s-1}\prod_{r=1}^{s-1}(1-F(\hat{c}_r^S))}{1-\sum_{s=1}^{t}F(\hat{c}_s^S)(1-\gamma)^{s-1}\prod_{r=1}^{s-1}(1-F(\hat{c}_r^S))}. \tag{A.36}$$

As with naïveté, the hazard rate is higher in period t than in $t+1$ if the memory-limitation effect dominates the option-value effect, which holds true whenever the memory-limitation effect is sufficiently strong, which holds true when the inattention probability γ is sufficiently high.

Consider next the case without a deadline/an infinitely long deadline, $T \to \infty$. Since the environment is time-invariant, the option value is time-invariant, too. The option value solves

$$0 = \int_0^{y-\delta\bar{V}^S}(y-\delta\bar{V}^S-c)dF(c) - (1-\delta(1-\gamma))\bar{V}^S. \tag{A.37}$$

By the Intermediate Value Theorem and monotonicity, a unique solution $\bar{V}^S$ exists and $\bar{V}^S \in (0, y)$. The option-value effect is hence absent and the memory-limitation effect always dominates, which is why the hazard rate is decreasing over time. Note that this is qualitatively the same result as we obtained in case of naïveté, but the option value in case of sophistication, $\bar{V}^S$, is different than the one with naïveté, $\bar{V}$.

We now show that – in accordance with the case of naïveté – the hazard rate depends negatively on the absolute value of time, holding the time distance to the deadline fixed. Note that by equations (A.33), (A.35), and $V_{T+1} = 0$ it holds that

$$V_t^S\big|_{T=T'} = V_{t+s}^S\big|_{T=T'+s} \quad \text{for all } T' \in \mathbb{N} \text{ and } s \in \mathbb{N}. \tag{A.38}$$

[6] To be precise, $\hat{c}_1^S = y-(1-\gamma)\delta\int_0^y(y-c)dF(c)$ while $\hat{c}_1 = y-\delta\int_0^y(y-c)dF(c)$, such that $h_1^S = F(\hat{c}_1^S) > F(\hat{c}_1) = h_1$.

We can hence apply the formal arguments we used for the case with a naive agent to clarify that

$$h_t^S\big|_{T=T'} > h_{t+s}^S\big|_{T=T'+s} \quad \text{for all } \ T' \in \mathbb{N} \ \text{ and } \ s \in \mathbb{N}. \tag{A.39}$$

We finally show that – as in case with naïveté – the within-deadline performance rate can be decreasing in the length of the deadline. It is sufficient to consider the case where $y \to \bar{k}$ and the deadline is either $T' = 1$ or $T'' = 2$. With the deadline $T' = 1$, the within-deadline performance rate is

$$p(T') = h_1^S\big|_{T'=1} = F(y) = 1. \tag{A.40}$$

In contrast, with the deadline $T'' = 2$ the within-deadline performance rate is

$$p(T'') = 1 - \left(1 - h_1^S\big|_{T''=2}\right)\left(1 - h_2^S\big|_{T''=2}\right) = 1 - \left(1 - h_1^S\big|_{T''=2}\right)\gamma. \tag{A.41}$$

Note that since the option value $V_2^S > 0$ for all $\gamma \in (0,1)$, it holds that $\hat{c}_1^S\big|_{T''=2} < y$ and thus that $h_1^S\big|_{T''=2} = F\left(\hat{c}_1^S\big|_{T''=2}\right) < 1$ and so $p(T'') < 1$. The within-deadline performance rate is hence higher with the short than with the long deadline, $p(T') > p(T'')$. As in case of a naïveté, the long deadline then never catches up the short deadline in terms of task performance rates, i.e., the cumulative performance rates are permanently lower with the long than with the short deadline.

B.3 Present Bias

Suppose now that the agent has time-inconsistent, present-biased preferences as captured by the well-known $\beta - \delta$ approach (e.g., Laibson, 1997), where the present bias parameter β is between zero and one. We further let the agent receive the reward Y with a delay of $\ell > 0$ periods after the task is performed, so that $y = \delta^\ell Y$ in the benchmark case of exponential discounting. Note that having delayed rewards is realistic in our scenario, as well as in many other real-world settings.

We first discuss the case where the agent is naive concerning her present bias. The agent then uses the same expected payoffs as an agent without present bias to calculate her optimal strategy, despite the fact that her true expected payoffs are different. Performing the task in period t yields a discounted payoff of $\beta\delta^\ell Y - c_t$, whereas not doing so yields $\beta\delta V_{t+1}$. Accordingly, the agent optimally performs the task in period t if and only if

$$c_t \leq \hat{c}_t^B := \beta\left(\delta^\ell Y - \delta V_{t+1}\right), \tag{A.42}$$

where the superscript B denotes the case of biased preferences we consider here. Comparing the cost thresholds in (A.42) and (1), we see that a present-biased agent uses a lower cost threshold than a time-consistent one. Except for inducing lower cost thresholds, however, the analysis is not affected by the present bias. In particular, if a present-biased agent has no (further) memory limitations (i.e., if $\gamma = 0$), hazard rates are still increasing over time and only depend on the time left until the deadline but not on the absolute value of time.[7]

We next consider the case of a sophisticated present-biased agent. The sophisticated agent decides to perform the task in period t if and only if

$$c_t \leq \hat{c}_t^{SB} := \beta \left(\delta^\ell Y - \delta V_{t+1}^{SB} \right), \tag{A.43}$$

where V_{t+1}^{SB} is the option value of the sophisticated biased agent. Except for using a different (namely the correct) option value, the sophisticated agent thus uses the same decision rule as a naive agent, cf. (A.42). The remaining proofs are – except for the adjustment of the agent's decision rule – as in the case of a time-consistent agent.

B.4 Overoptimism

We next briefly analyze what happens when the agent is overoptimistic regarding the costs of task performance, in the sense that she overestimates the likelihood of low cost realizations and underestimates the likelihood of high cost realizations. Formally, an overoptimistic agent bases her decisions on a perceived cost distribution F^o, while her costs are actually drawn from distribution F, where $F^o(c) > F(c)$ for all $c \in (0, \bar{k})$, $F^o(0) = F(0) = 0$, and $F^o(\bar{k}) = F(\bar{k}) = 1$. The overoptimistic agent performs the task in period t if and only if

$$c_t \leq \hat{c}_t^o = y - \delta V_t^o. \tag{A.44}$$

In period T, the perceived option value of an optimistic agent is higher than the option value of an agent with realistic beliefs:

$$V_T^o = \int_0^{\hat{c}_T^o = y} (y - c) dF^o(c) > \int_0^{\hat{c}_T = y} (y - c) dF(c) = V_T. \tag{A.45}$$

Since

$$V_t^o = \int_0^{\hat{c}_t^o} (y - c) dF^o(c) + (1 - F^o(\hat{c}_t^o)) \, \delta V_{t+1}^o > \int_0^{\hat{c}_t} (y - c) dF(c) + (1 - F(\hat{c}_t)) \, \delta V_{t+1} = V_t, \tag{A.46}$$

the same holds true in period $T - 1$ and, by induction, also in all previous periods. Hence, overoptimism causes the agent to perceive the option value to be strictly higher than its true value in all periods $t \leq T$. From (1) and (A.44), it then follows that an overoptimistic agent uses lower cost thresholds in all $t \leq T - 1$, compared to an agent with realistic beliefs and no memory limitations. Similar to a present-biased agent, the overoptimistic one therefore exhibits, ceteris paribus, a higher likelihood to postpone the task in any given period. Beyond these quantitative differences, however, the results remain qualitatively unchanged relative to the no-memory-limitations case.

[7] For an analysis capturing possible interactions between present-bias and memory limitations see Ericson (2017).

B.5 Agent's Well-being

The agent's expected payoff at the beginning of the decision problem is measured by V_1. In case of no memory limitations, the agent's expected payoff V_1 is increasing in the length of the deadline T. This is directly implied by (A.3) and (A.10). An agent with no memory limitations thus benefits from the additional freedom a longer deadline provides.

Under memory limitations, however, the link between the length of the deadline and the agent's expected payoff becomes ambiguous. In fact, with memory limitations, the agent's expected payoff might suffer from a longer deadline. To see this, it is useful to consider the case where inattention probability γ is large, $\gamma \to 1$. The agent's expected payoff is then

$$V_1 = \int_0^{\hat{c}_1} (y - c) dF(c). \tag{A.47}$$

Because the threshold $\hat{c}_1$ falls short of the reward y and decreases in the length of the deadline T (since by (A.3) and (A.10) the option value V_2 is decreasing in T), we directly see from (A.47) that the agent's expected payoff V_1 is decreasing in the length of the deadline T.

Agents with memory limitations might therefore be better off under a shorter deadline. The intuition for this result is that a longer deadline induces the agent to complete the task only for relatively low cost realizations. However, applying rather low cost thresholds together with limited memory makes the agent's task performance rather unlikely, such that the agent is rather unlikely to enjoy the rewards for carrying out the task. Put differently, a longer deadline may provoke that an agent with limited memory lowers her thresholds for which she is willing to perform the task so much that she harms herself. In general, one can directly show (by use of simple continuity arguments) that a longer deadline harms an agent if γ is sufficiently high.

The result that a longer deadline can harm an agent crucially depends on an agent's naïveté. If an agent is fully sophisticated – in the sense of fully aware of her memory limitations – she optimally adjusts her strategy and continues to benefit from longer deadlines. Sophistication is analyzed below in detail.

B.6 Recall

By continuity, our results are also robust to the case where – after becoming inattentive – the agent may later return to processing the task with a small positive probability. More generally, one can show that a model with positive probability of 'recall' is isomorphic to the model with 'zero recall' and time-dependent probabilities of becoming inattentive $(\gamma_1, \ldots, \gamma_T)$.

B.7 Additional Costs

In our experiment, in addition to the costs for arranging a check-up, there are also costs for attending an appointment at the clinic. The model can be readily extended to capture these type of costs. The modification of the model is rather simple: we only have to reinterpret the task-performance cost c as the sum of the costs of arranging and the expected costs of attending the appointment. Note further that the primary task in the experiment – the action that is tied to the deadline – is arranging an appointment.

C Complementary Material

C.1 Reminder Postcard and Timing of Waves

Figure I presents the layout of a reminder postcard. The figure displays an example featuring a deadline and a free dental cleaning as reward. The text on the backside of the postcard (ride side of Figure I) can be translated as follows: '*Please call us to arrange an appointment for your next check-up. If you contact us by DD / MM, you receive—if desired—a free professional dental cleaning worth 70 euros. Please inform us when arranging the appointment.*' The corresponding deadline in treatments D1 and D3 was filled into the blank space on the postcard. The text on the postcards that featured no deadline and / or small or no explicit rewards were adapted accordingly. The front side of the postcard (left part of Figure I) was held constant across treatments.

Timing. During the experiment, the dentist sent out reminders every second Friday, implying that the postcards are delivered to participants on Saturday or Monday, at the latest. The deadlines of the D1 and the D3 treatment were set for the Friday one and three weeks after the sending day, respectively. Together with the biweekly sending waves, this implies that the deadline date of the D3 treatment from wave w coincides with the deadline from the D1 treatment in wave $w + 1$.

Figure I: Reminder Postcard

C.2 Complementary Results

Table I: Duration Analysis (Main Experiment)

	(1)	(2)	(3)
D1	1.135	1.157	1.158
	[0.103]	[0.066]	[0.074]
D3	1.180	1.215	1.223
	[0.028]	[0.012]	[0.012]
Small Incent		1.085	1.086
		[0.336]	[0.346]
Large Incent		1.429	1.441
		[0.000]	[0.000]
ρ (shape parameter)			0.565
			[0.000]
Controls	–	Yes	Yes

Notes: The table presents hazard ratios estimated with Cox (Columns 1 and 2) proportional hazards models (N = 2,661) with the structure $h_t = h_{0,t} \exp(\beta_0 + \beta_1 \text{D1} + \beta_2 \text{D3})$. While Cox proportional hazards model leaves the baseline hazard function $h_{0,t}$ unspecified (estimating it non-parametrically). Column (3) reports the results from the Weibull model, which is based on a parametric specification with $h_{0,t} = \rho t^{\rho-1}$, where ρ is the estimated shape parameter of the Weibull distribution. The Weibull model can thus accommodate a baseline hazard that is either decreasing ($0 < \rho < 1$), constant ($\rho = 1$), or increasing ($\rho > 1$). Specifications (2) and (3) include dummies for different incentive levels, controls for individual characteristics, and dummies absorbing wave specific effects. p-values, based on robust standard errors, clustered at the household level, are reported in brackets.

Table II: Duration Analysis with Time-Varying Treatment Effects (Main Experiment)

	(1)	(2)	(3)
$D1^{w1}$	3.506	3.574	3.312
	[0.000]	[0.000]	[0.000]
$D1^{w2}$	1.266	1.299	1.237
	[0.193]	[0.154]	[0.290]
$D1^{w3}$	0.819	0.835	0.615
	[0.375]	[0.426]	[0.038]
$D1^{w4+}$	0.672	0.675	0.726
	[0.000]	[0.000]	[0.002]
$D3^{w1}$	2.588	2.632	2.400
	[0.000]	[0.000]	[0.000]
$D3^{w2}$	1.667	1.717	1.693
	[0.003]	[0.002]	[0.005]
$D3^{w3}$	2.444	2.505	1.805
	[0.000]	[0.000]	[0.000]
$D3^{w4+}$	0.678	0.699	0.752
	[0.000]	[0.001]	[0.011]
Controls	–	Yes^a	Yes^b
Post-Estimation Tests:			
$D3^{w1} = D3^{w2}$	0.026	0.030	0.114
$D3^{w2} = D3^{w3}$	0.062	0.063	0.778
$D1^{w1} = D3^{w3}$	0.041	0.046	0.001

Notes: The table presents hazard ratios estimated with Weibull proportional hazards model (N = 2,661). Specification (2) includes dummies for different incentive levels, controls for individual characteristics, and dummies absorbing wave specific effects. Specification (3) adds further dummies that account for week×year specific effects. p-values – based on robust standard errors, clustered at the household level – are reported in brackets. The lower part of the table reports p-values from Wald tests.

Table III: Duration Analysis with Time-Varying Treatment Effects (Main Experiment)

	(1)	(2)	(3)	(4)	(5)	(6)
	$Z = 1$ (last day)			$Z = 2$ (last two day)		
$\text{D1}^{\text{w1-w/o-last}Z\text{days}}$	3.493	3.567	3.509	3.363	3.446	3.695
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
$\text{D1}^{\text{w1-last}Z\text{days}}$	3.576	3.623	2.770	3.684	3.732	2.941
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
D1^{w2}	1.266	1.299	1.239	1.261	1.294	1.241
	[0.194]	[0.155]	[0.286]	[0.202]	[0.161]	[0.284]
D1^{w3}	0.819	0.835	0.612	0.818	0.833	0.610
	[0.375]	[0.426]	[0.036]	[0.371]	[0.421]	[0.035]
$\text{D1}^{\text{w4+}}$	0.672	0.675	0.723	0.675	0.677	0.721
	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.001]
D3^{w1}	2.587	2.633	2.444	2.555	2.603	2.470
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
D3^{w2}	1.667	1.717	1.696	1.660	1.710	1.698
	[0.003]	[0.002]	[0.005]	[0.003]	[0.002]	[0.005]
$\text{D3}^{\text{w3-w/o-last}Z\text{days}}$	2.631	2.709	1.946	2.319	2.408	1.722
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.005]
$\text{D3}^{\text{w3-last}Z\text{days}}$	1.600	1.608	1.140	2.641	2.652	1.900
	[0.216]	[0.213]	[0.734]	[0.000]	[0.000]	[0.004]
$\text{D3}^{\text{w4+}}$	0.678	0.699	0.749	0.680	0.701	0.748
	[0.000]	[0.001]	[0.010]	[0.001]	[0.002]	[0.009]
Controls	–	Yes^a	Yes^b	–	Yes^a	Yes^b
Post-Estimation Tests: Comparing last Z days						
$\text{D1}^{\text{w1-last}Z\text{days}} = \text{D3}^{\text{w3-last}Z\text{days}}$	0.063	0.061	0.040	0.198	0.185	0.096

Notes: The table presents hazard ratios estimated with Weibull proportional hazards model ($N = 2,661$). Specifications (2) and (4) includes dummies for different incentive levels, controls for individual characteristics, and dummies absorbing wave specific effects. Specifications (3) and (5) adds further dummies that account for week×year specific effects. p-values – based on robust standard errors, clustered at the household level – are reported in brackets.

Table IV: Treatment Effects on Period-t Performance Rates (Follow-up Experiment)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	7 Days		21 Days		42 Days		70 Days		100 Days	
D1	0.111**	0.107**	0.050	0.046	0.064	0.066	0.054	0.056	0.038	0.039
	(0.046)	(0.044)	(0.052)	(0.051)	(0.055)	(0.053)	(0.055)	(0.053)	(0.056)	(0.054)
D3	0.053	0.052	0.140***	0.137***	0.108*	0.100*	0.100*	0.087*	0.091	0.074
	(0.041)	(0.041)	(0.052)	(0.050)	(0.055)	(0.052)	(0.056)	(0.053)	(0.057)	(0.054)
D6	-0.033	-0.029	-0.050	-0.045	0.034	0.037	0.051	0.050	0.023	0.024
	(0.036)	(0.036)	(0.046)	(0.046)	(0.056)	(0.053)	(0.058)	(0.055)	(0.059)	(0.056)
D10	-0.047	-0.045	-0.047	-0.050	-0.038	-0.035	0.019	0.019	0.022	0.021
	(0.035)	(0.036)	(0.048)	(0.049)	(0.051)	(0.052)	(0.055)	(0.054)	(0.056)	(0.054)
EoY	0.000	0.027	-0.009	0.018	-0.011	0.026	0.041	0.073	0.029	0.055
	(0.046)	(0.046)	(0.059)	(0.058)	(0.061)	(0.061)	(0.063)	(0.063)	(0.064)	(0.065)
Constant	0.131***	-0.097	0.240***	-0.055	0.297***	0.073	0.354***	0.199	0.400***	0.197
	(0.027)	(0.113)	(0.035)	(0.174)	(0.038)	(0.186)	(0.038)	(0.188)	(0.040)	(0.194)
Post-Estimation Tests: (p-value)										
D1=D3	0.207	0.220	0.090	0.073	0.425	0.519	0.424	0.571	0.355	0.516
D1=D6	0.001	0.002	0.040	0.059	0.601	0.584	0.954	0.910	0.806	0.802
D1=D10	0.000	0.000	0.048	0.053	0.051	0.054	0.530	0.500	0.771	0.748
D1=EoY	0.035	0.108	0.327	0.625	0.231	0.514	0.846	0.791	0.889	0.803
D3=D6	0.025	0.037	0.000	0.000	0.196	0.240	0.409	0.510	0.258	0.382
D3=D10	0.009	0.011	0.000	0.000	0.009	0.014	0.154	0.219	0.225	0.328
D3=EoY	0.286	0.609	0.016	0.046	0.060	0.223	0.375	0.826	0.355	0.773
D6=D10	0.667	0.638	0.953	0.914	0.178	0.182	0.592	0.598	0.978	0.958
D6=EoY	0.447	0.201	0.476	0.266	0.476	0.870	0.890	0.722	0.939	0.649
Controls	–	Yes	–	Yes	–	Yes	–	Yes		

Notes: The table presents LPM estimates of the follow-up experiment treatment effects on the cumulative response rate (the probability of calling) within 7, 21, 42, 70 and 100 days, respectively ($N = 927$). The lower part reports the p-values from Wald tests. Every second specification includes individual control variables and dummies absorbing wave specific effects. Robust standard errors, clustered at the household level, are reported in parentheses. ***, **, * indicates significance at the 1%-, 5%-, 10%-level, respectively.

Table V: Duration Analysis (Follow-up Experiment)

	(1)	(2)	(3)
D1	1.201	1.205	1.203
	[0.308]	[0.305]	[0.327]
D3	1.391	1.336	1.350
	[0.056]	[0.089]	[0.093]
D6	1.078	1.022	1.020
	[0.672]	[0.904]	[0.915]
D10	1.033	0.998	0.998
	[0.850]	[0.991]	[0.993]
EoY	1.079	1.205	1.215
	[0.702]	[0.383]	[0.382]
ρ (shape parameter)			0.563
			[0.000]
Controls	–	Yes	Yes
Post-Estimation Tests: (p-values)			
D1=D3	0.395	0.549	0.523
D1=D6	0.545	0.362	0.381
D1=D10	0.374	0.276	0.301
D1=EoY	0.603	0.999	0.965
D3=D6	0.134	0.116	0.118
D3=D10	0.072	0.075	0.079
D3=EoY	0.206	0.616	0.626
D6=D10	0.801	0.892	0.906
D6=EoY	0.998	0.437	0.431
D10=EoY	0.827	0.371	0.374

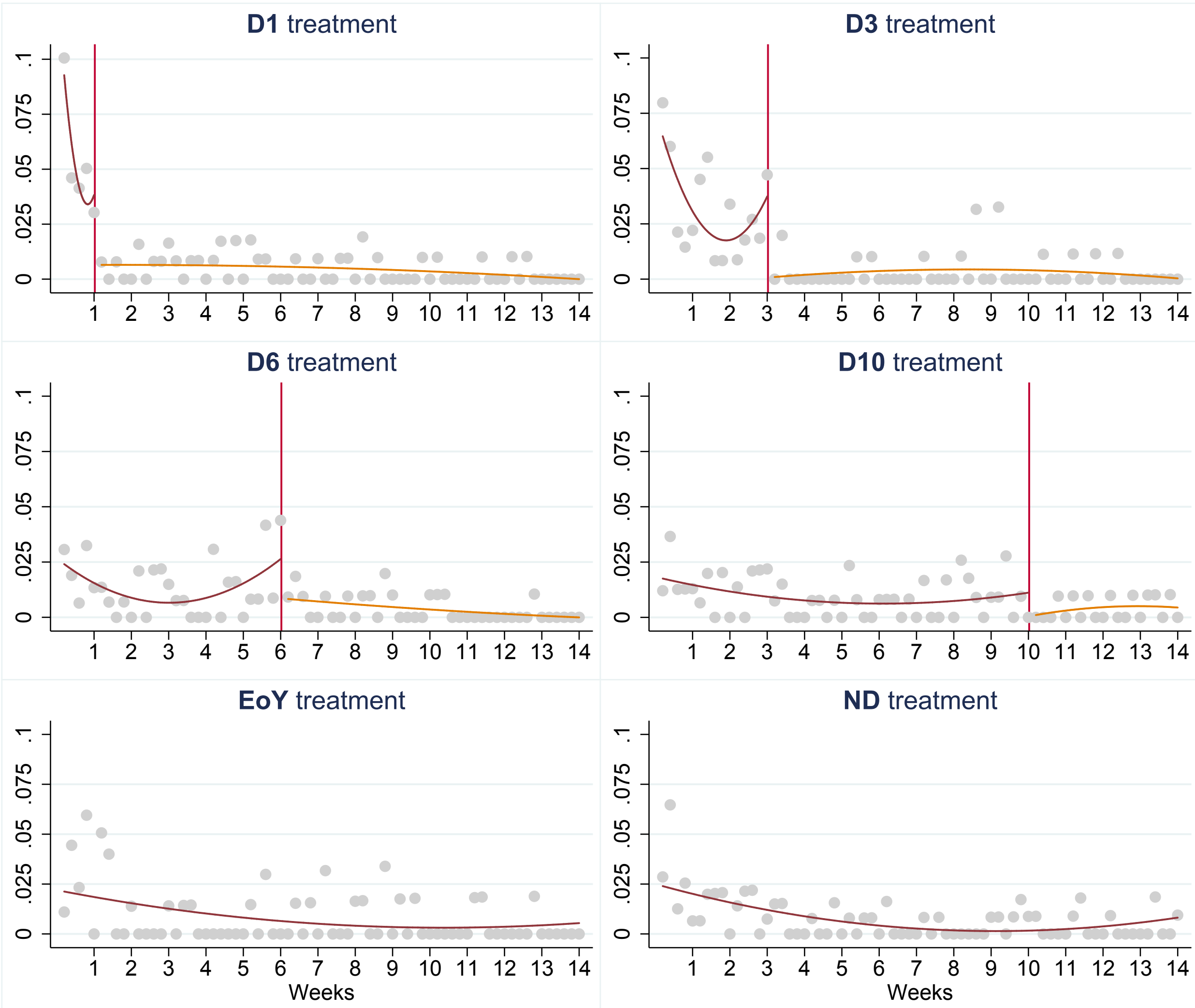
Notes: The table presents hazard ratios estimated with Cox (columns 1 and 2) and Weibull (column 3) proportional hazards model ($N = 927$). Specifications (2) and (3) include dummies for different incentive levels, controls for individual characteristics, and dummies absorbing wave specific effects. p-values—based on robust standard errors, clustered at the household level—are reported in brackets. ρ is the estimated shape parameter of the Weibull distribution, with $\rho < 1$ indicating a decreasing hazard rate. The lower part of the table reports p-values from Wald tests.

Table VI: Duration Analysis with Time-Varying Treatment Effects (Follow-up Experiment)

	Estimation Output – Part 1 (1)	(2)	(3)		*Estimation Output – Part 2* (1)	(2)	(3)
$\mathrm{D1}^{\mathrm{w1}}$	3.731	3.691	3.663	$\mathrm{D6}^{\mathrm{w1}}$	1.394	1.334	1.322
	[0.000]	[0.000]	[0.000]		[0.308]	[0.385]	[0.424]
$\mathrm{D1}^{\mathrm{w2}}$	0.329	0.328	0.205	$\mathrm{D6}^{\mathrm{w2}}$	0.580	0.538	0.341
	[0.124]	[0.124]	[0.030]		[0.297]	[0.238]	[0.044]
$\mathrm{D1}^{\mathrm{w3}}$	1.155	1.161	0.582	$\mathrm{D6}^{\mathrm{w3}}$	1.882	1.749	0.878
	[0.737]	[0.731]	[0.216]		[0.060]	[0.103]	[0.706]
$\mathrm{D1}^{\mathrm{w4\text{-}5}}$	0.922	0.920	0.618	$\mathrm{D6}^{\mathrm{w4\text{-}5}}$	1.049	0.974	0.667
	[0.829]	[0.822]	[0.198]		[0.887]	[0.939]	[0.220]
$\mathrm{D1}^{\mathrm{w6}}$	1.026	1.025	0.906	$\mathrm{D6}^{\mathrm{w6}}$	3.110	2.877	2.627
	[0.960]	[0.962]	[0.850]		[0.011]	[0.015]	[0.020]
$\mathrm{D1}^{\mathrm{w7\text{-}9}}$	0.678	0.685	0.884	$\mathrm{D6}^{\mathrm{w7\text{-}9}}$	1.074	1.024	1.346
	[0.328]	[0.343]	[0.759]		[0.836]	[0.946]	[0.401]
$\mathrm{D1}^{\mathrm{w10}}$	0.324	0.333	0.591	$\mathrm{D6}^{\mathrm{w10}}$	0.333	0.322	0.581
	[0.263]	[0.274]	[0.607]		[0.273]	[0.263]	[0.599]
$\mathrm{D1}^{\mathrm{w11+}}$	0.419	0.422	1.238	$\mathrm{D6}^{\mathrm{w11+}}$	0.361	0.345	1.044
	[0.042]	[0.045]	[0.661]		[0.030]	[0.024]	[0.936]
$\mathrm{D3}^{\mathrm{w1}}$	2.763	2.579	2.541	$\mathrm{D10}^{\mathrm{w1}}$	1.185	1.142	1.126
	[0.000]	[0.001]	[0.002]		[0.624]	[0.706]	[0.743]
$\mathrm{D3}^{\mathrm{w2}}$	3.219	2.968	1.903	$\mathrm{D10}^{\mathrm{w2}}$	0.987	0.934	0.579
	[0.000]	[0.000]	[0.036]		[0.975]	[0.864]	[0.182]
$\mathrm{D3}^{\mathrm{w3}}$	2.814	2.668	1.328	$\mathrm{D10}^{\mathrm{w3}}$	1.859	1.782	0.900
	[0.000]	[0.001]	[0.359]		[0.060]	[0.086]	[0.758]
$\mathrm{D3}^{\mathrm{w4\text{-}5}}$	0.271	0.266	0.172	$\mathrm{D10}^{\mathrm{w4\text{-}5}}$	0.616	0.588	0.404
	[0.068]	[0.063]	[0.014]		[0.249]	[0.208]	[0.030]
$\mathrm{D3}^{\mathrm{w6}}$	0.576	0.558	0.497	$\mathrm{D10}^{\mathrm{w6}}$	1.134	1.072	0.953
	[0.442]	[0.419]	[0.334]		[0.784]	[0.881]	[0.920]
$\mathrm{D3}^{\mathrm{w7\text{-}9}}$	0.531	0.517	0.692	$\mathrm{D10}^{\mathrm{w7\text{-}9}}$	1.217	1.175	1.539
	[0.169]	[0.147]	[0.415]		[0.506]	[0.593]	[0.166]
$\mathrm{D3}^{\mathrm{w10}}$	1.121	1.124	2.055	$\mathrm{D10}^{\mathrm{w10}}$	1.546	1.536	2.820
	[0.848]	[0.846]	[0.260]		[0.426]	[0.434]	[0.074]
$\mathrm{D3}^{\mathrm{w11+}}$	0.553	0.543	1.747	$\mathrm{D10}^{\mathrm{w11+}}$	0.616	0.610	1.986
	[0.132]	[0.128]	[0.218]		[0.216]	[0.203]	[0.132]
Controls	–	Yes^{a}	Yes^{b}				
Post-Estimation Tests: (p-values)							
$\mathrm{D1}^{\mathrm{w1}} = \mathrm{D3}^{\mathrm{w3}}$	0.415	0.341	0.007				
$\mathrm{D1}^{\mathrm{w1}} = \mathrm{D6}^{\mathrm{w6}}$	0.709	0.599	0.488				
$\mathrm{D1}^{\mathrm{w1}} = \mathrm{D10}^{\mathrm{w10}}$	0.142	0.143	0.689				
$\mathrm{D3}^{\mathrm{w3}} = \mathrm{D6}^{\mathrm{w6}}$	0.842	0.878	0.160				
$\mathrm{D3}^{\mathrm{w3}} = \mathrm{D10}^{\mathrm{w10}}$	0.321	0.362	0.255				
$\mathrm{D6}^{\mathrm{w6}} = \mathrm{D10}^{\mathrm{w10}}$	0.306	0.353	0.918				

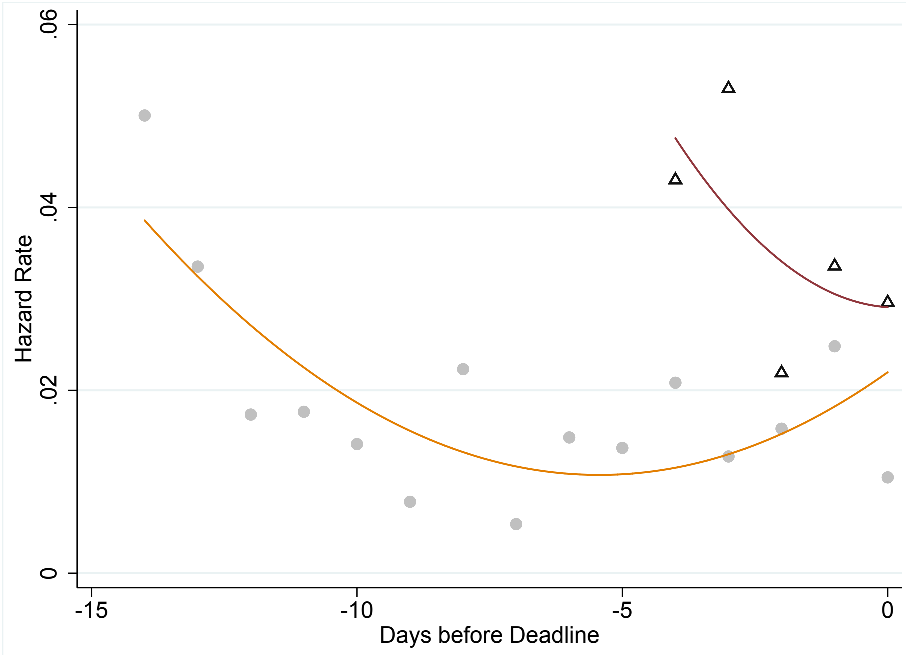
Notes: The table presents hazard ratios estimated with Weibull proportional hazards model ($N = 927$). Specification (2) includes dummies for different incentive levels, controls for individual characteristics, and dummies absorbing wave specific effects. Specification (3) adds further dummies that account for week×year specific effects. All specifications control for the (time-invariant and statistically insignificant) difference between the No-Deadline and End-of-Year Deadline treatment. p-values – based on robust standard errors, clustered at the household level – are reported in brackets. The lower part of the table reports p-values from Wald tests.

Figure II: Daily Hazard Rates (Follow-up Experiment)



Notes: The figure displays daily hazard rates for all treatments of the follow-up experiment. The graph covers a period of 14 weeks after sending the mailings.

Figure III: Daily Hazard Rates – Re-scaled Time Frame (Main Experiment)

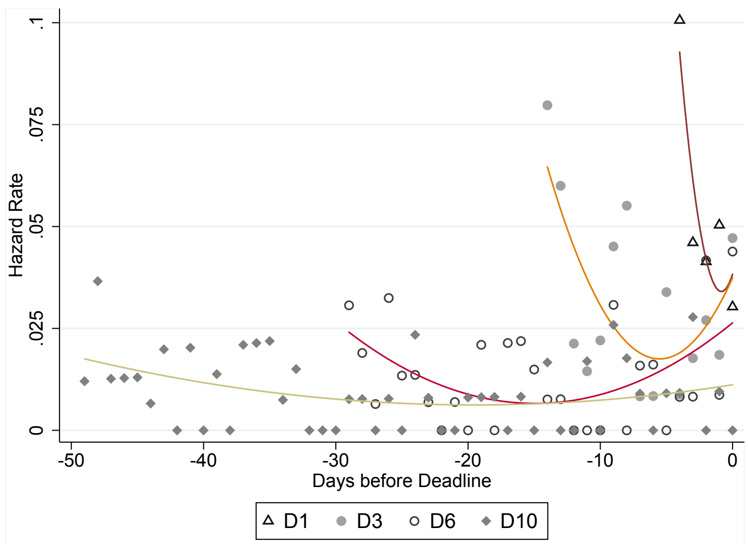


Notes: The figure plots daily hazard rates for the pre-deadline period, comparing the D1 (indicated by the triangles, △) and the D3 treatment (gray dots, ●).

In addition to the weekly differentiation of treatment effects reported in Table VI, we also studied treatment differences during the pre-deadline day(s), in the spirit of the analysis reported in Figure III and Table III. Consistent with the evidence from Figure IV below, the analysis shows that the hazard rate for the last days before the ten-week deadline tends to be significantly lower than the hazards immediately before the shorter deadlines.

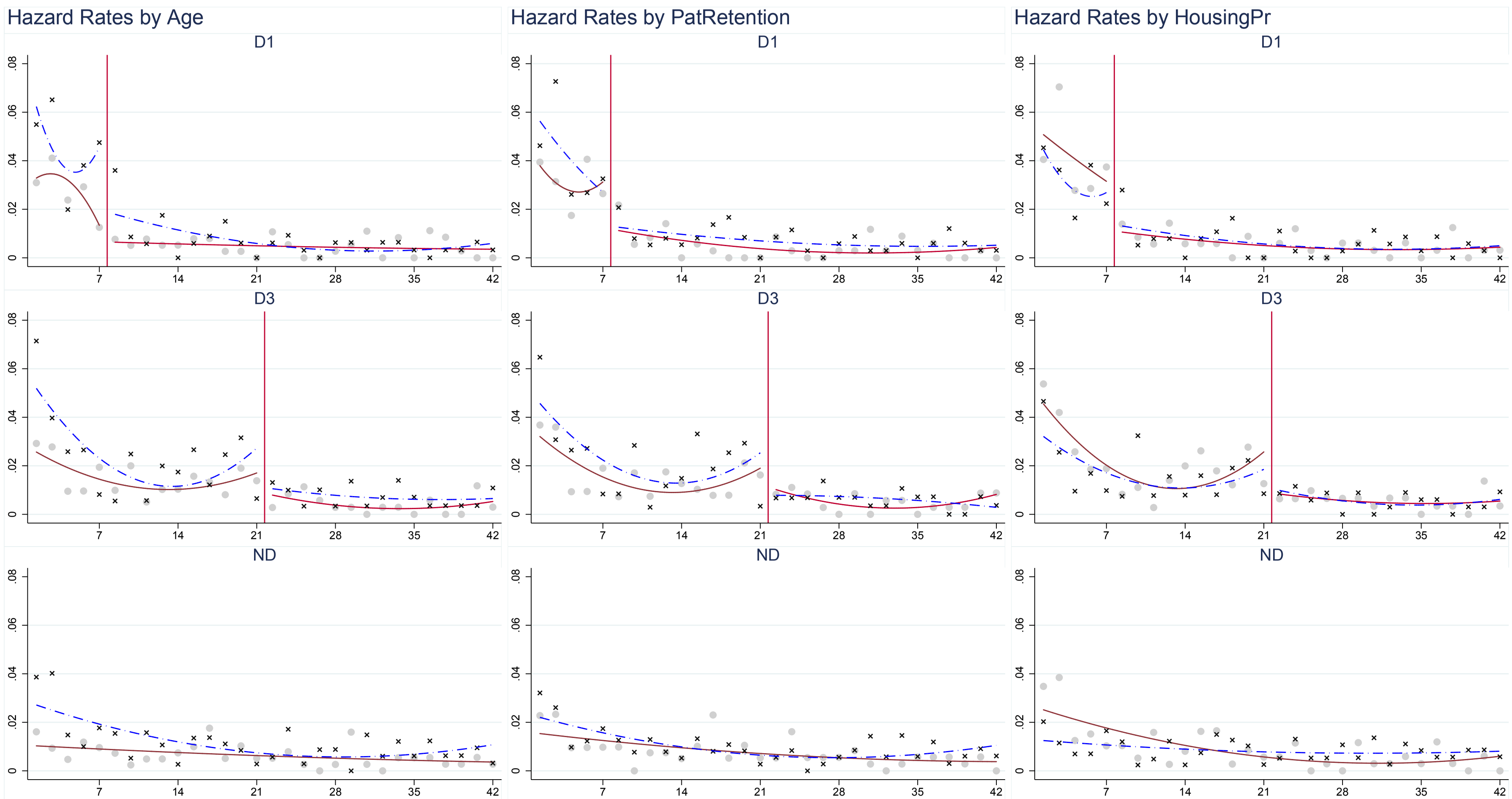
More specifically, hazard rates during the *last three days* before the ten-week deadline are significantly below the values for the corresponding pre-deadline days in the D1, D3, and D6 treatment, respectively (p-values ranging from 0.052 to 0.072). When considering only the last *two days* before the deadlines, differences are still sizable, but lose statistical significance (p-values around 0.15). For the day immediately before the deadline, the hazard rate in the D10 treatment is exactly zero (see Figure II), whereas hazard rates for the pre-deadline day are significantly positive in all other treatments. Further details on the estimation results are available upon request.

Figure IV: Daily Hazard Rates (Follow-up Experiment)



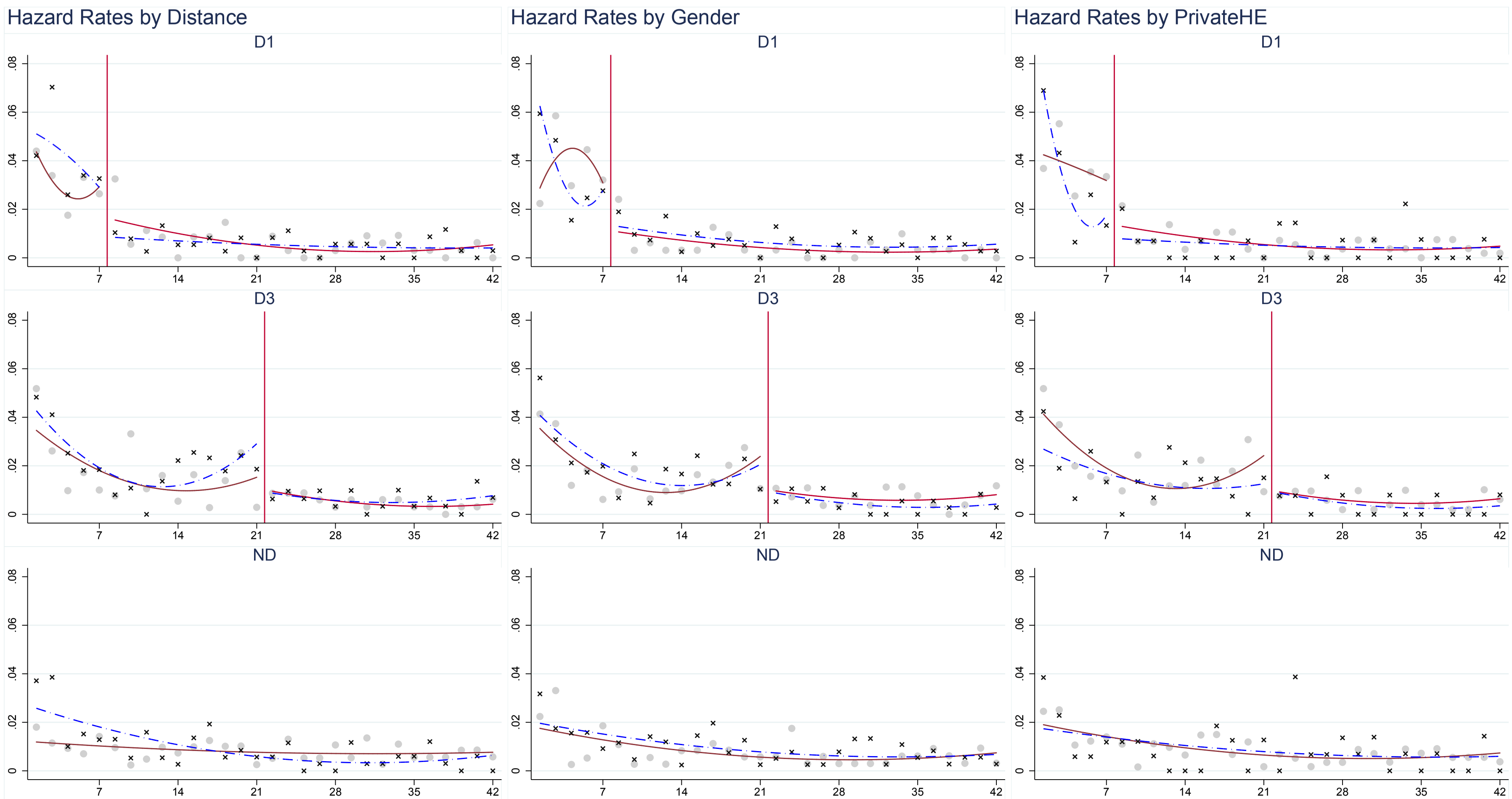
Notes: The figure displays daily hazard rates – for the period before the deadline – for the one-, three-, six-, and 10-weeks deadline treatment of the follow-up experiment.

Figure V: Hazard Rates for Different Subgroups (1)



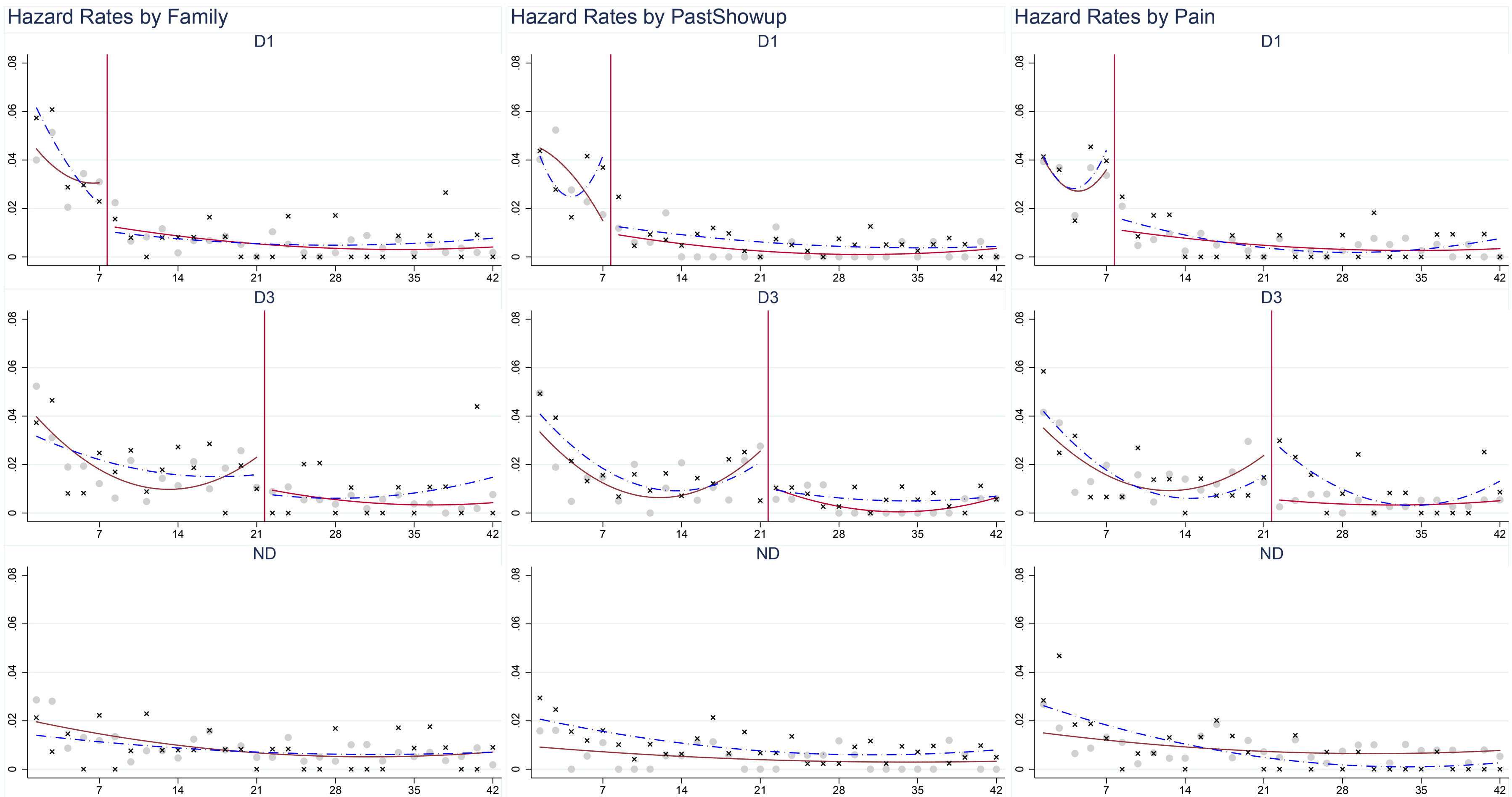
Notes: The figure displays daily hazard rates for the D1, D3 and ND treatment over a period of six weeks, splitting the sample by Age, Patient retention and Housing price (see Table A.1 for variable descriptions): gray dots ● (and solid, red lines) depict hazard rates for samples with below median Age, Patient retention and Housing price, respectively. Black × (and dashed, blue lines) indicate hazard rates for the corresponding above median samples.

Figure VI: Hazard Rates for Different Subgroups (2)



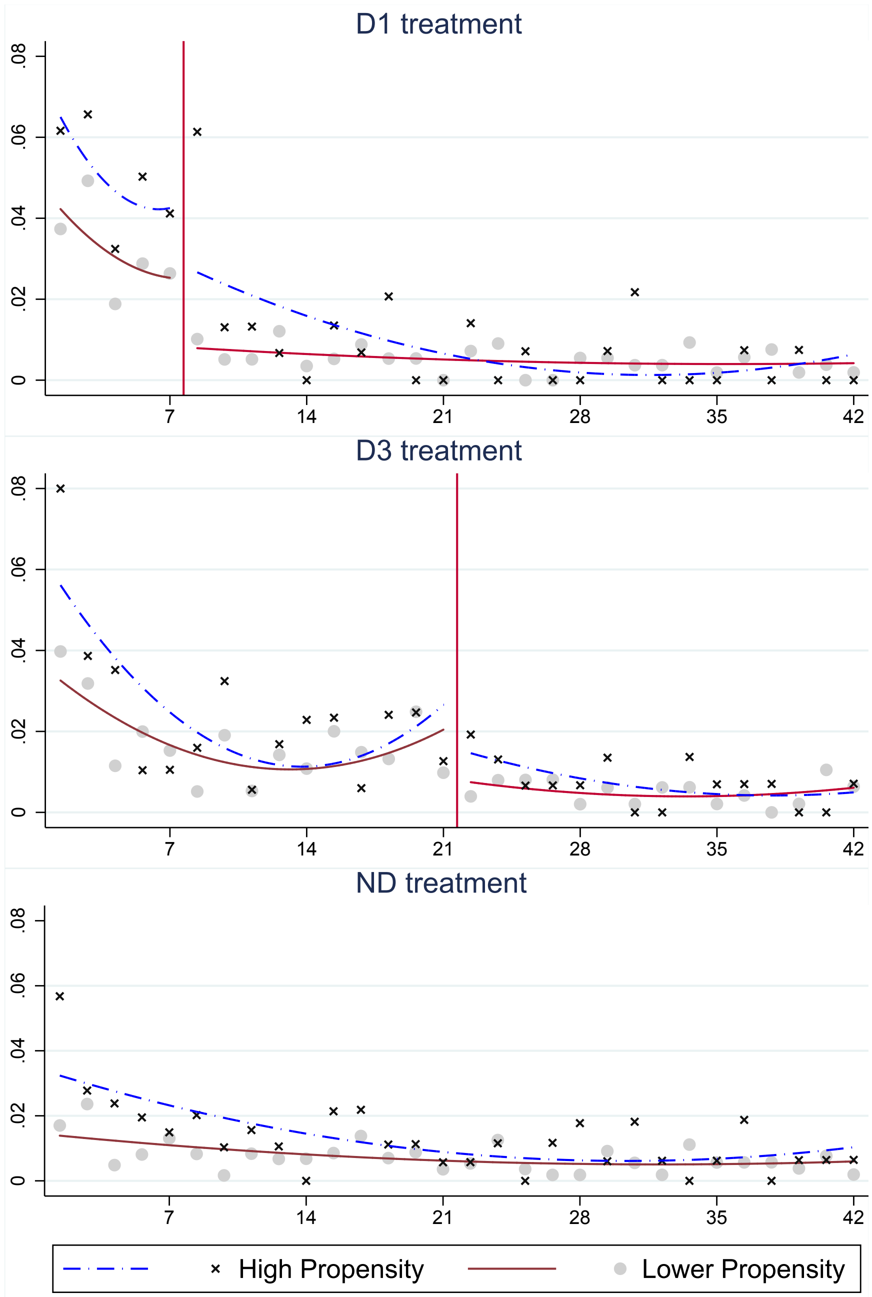
Notes: The figure displays daily hazard rates for the D1, D3 and ND treatment over a period of six weeks, splitting the sample by Distance, Gender, and Private health insurance status (see Table A.1 for variable descriptions): gray dots ● (and solid, red lines) depict hazard rates for samples with below median distance, for males (44% of sample) and for patients in the public heath insurance system (80%), respectively. Black × (and dashed, blue lines) indicate hazard rates for above median distance, for females, and for patients covered by private health insurance, respectively.

Figure VII: Hazard Rates for Different Subgroups (3)



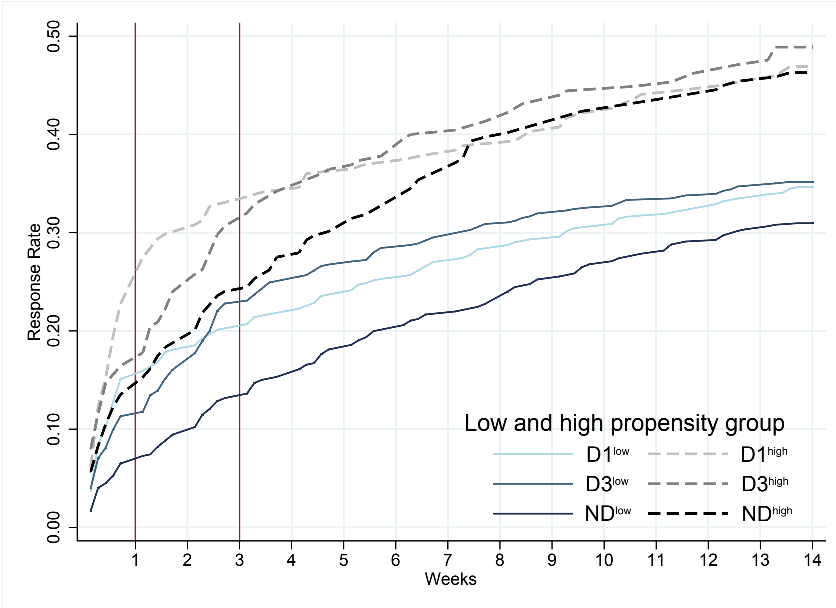
Notes: The figure displays daily hazard rates for the D1, D3 and ND treatment over a period of six weeks, splitting the sample by 'Family', 'Past showup'$^{\flat}$, and 'Pain'$^{\flat}$ (see Table A.1 for variable descriptions): gray dots ● (and solid, red lines) depict hazard rates for non-family patients (84% of sample), patients that did not have a check-up in the past year (28%) and patients that did not have a painful dental treatment (75%), respectively. Black × (and dashed, blue lines) indicate hazard rates for families, patients that did have at least one check-up and patients that did experience a painful dental treatment in the past, respectively. (The two variables labeled with ♭ are only available for 2,189 [Past showup] and 1,990 [Pain] observations, respectively.)

Figure VIII: Hazard Rates for Subgroups (Main Experiment)



Notes: The figure displays raw, daily hazard rates for the D1, D3 and ND treatment over a period of six weeks (42 days), splitting the sample among those with the highest (top 25%) and with lower predicted propensities to respond (see text for details).

Figure IX: Cumulative Response Rates for Subgroups (Main Experiment)



Notes: The figure presents the empirically observed cumulative response rates over 14 weeks, splitting the sample among those with the highest (top 25%) and with lower predicted propensities to respond (see text for details).

C.3 Simulations: Type Heterogeneity

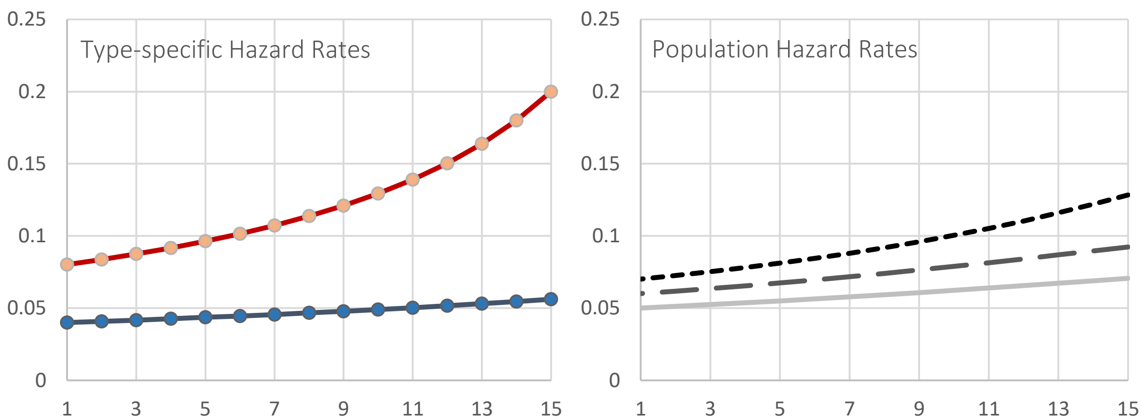
In this section, we summarize the results of numerical simulations that study how type heterogeneity and differential sorting can affect population hazards in our setting. Our analysis starts out from the observation that hazard rates on the first day of the intervention are not affected by differential sorting. Hence, day-1 hazard rates offer a straightforward proxy for the degree of heterogeneity in patients' responsiveness. We build on this metric to calibrate opportunity-cost distributions and hazard rates of different subgroups in our numerical simulations. In a first set of simulations, we consider mixtures of types with *linear cost distributions* $F(c)$ and first focus on differential sorting in the context of a three-week deadline, corresponding to the D3 treatment presented in the middle panel of Figure 4. (One arrives at similar results if the analysis departs from the observed day-1 hazards under the one-week deadline (D1 treatment); the latter treatment is discussed in more detail below.)

We started with uniform cost distributions that match the day-1 hazard rates observed for the subgroup with the highest response propensity ($= 0.08$) and the group with lower response propensities ($= 0.04$; see Figure VIII). Based on these distributions, we then compute type-specific as well as population hazards. The upper left panel of Figure X depicts the type-specific hazard rates for the high- and low-response type over the three weeks leading up to the D3 deadline (corresponding to 15 periods/working days in our context). In line with our model framework without memory limitations, hazard rates are increasing both for the high-response (red line) and low-response type (blue line).

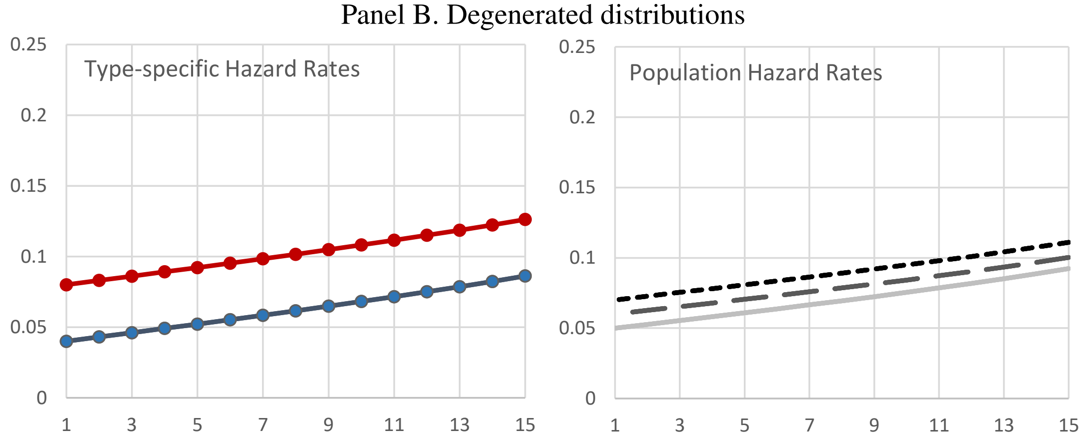
The top right panel of Figure X illustrates the resulting population hazards for different mixtures of these two types. The solid grey line considers the actual population shares underlying the heterogeneity analysis from Figure VIII, i.e., a population comprising 25% highly responsive patients and 75% with a lower response propensity. The dashed and dotted lines depict scenarios with alternative type mixtures, which assume a population share of high-response types of 50% (dashed grey line) and 75% (dotted black line), respectively.[8] For none of these hypothetical cases we find locally decreasing population hazards.

Figure X: Type specific and population hazard rates (3-week deadline)

Panel A. Linear cost distributions



Panel B. Degenerated distributions



Notes: The left panels illustrate type-specific hazard rates for a low- and a high-response type over 15 periods. The right panels plot the resulting population hazard rates, based on a 'low'/'high'-type mixture with 75/25 (solid grey line), 50/50 (dark grey, dashed line) and 25/75% (black, dotted line), respectively. In Panel A, the low-response type (blue line) and high-response type (red line) are based on uniform cost distributions with support [0,17.8] and [0,5.0], respectively. The reward is normalized to $y = 1$ and $\delta = 1$. The types in Panel B are based on degenerated cost distributions.

Next we considered mixtures of types with different linear cost distributions (e.g., triangular distributions). Starting from empirically plausible day-1 hazards, we failed to find parametrizations that would yield decreasing or U-shaped population hazards like the ones we observe in our experiments. In fact, for special cases

[8] The latter scenario is empirically implausible, as it inverts the starting point of our analysis (a 25/75 mix of high/low responsive types); the case is nevertheless useful to examine the boundaries of the simulations.

in which the cost distributions of different types belong to the same family of (e.g., linear) distributions with support of the individual distributions being subsets of $[0; y]$, one can show analytically that type heterogeneity never leads to locally decreasing population hazards.[9]

Moving beyond linear cost functions, we next turned to mixtures of types with a wider range of alternative distributions. We started off with a simple case based on two types with *degenerated cost functions* that imply type-specific hazard rates which again match the empirically observed day-1 hazards of 4% and 8% and move in parallel (see bottom left panel of Figure X). (Note that, relative to the case depicted in Panel A of Figure X, the weaker increase in hazards of the high type mitigates the 'upward pressure' on population hazards.) The bottom right panel of Figure X presents the resulting population hazards. The hazard rates are again inconsistent with the U-shaped hazard observed in the D3 treatment of our experiments; for all three scenarios with different population shares of the high- and low-response type, one obtains increasing hazards.

Starting from this example, we explored more extreme scenarios. In particular, we considered cases in which the more responsive type has significantly higher day-1 hazard rates. The intuition for doing so is straightforward: in order to generate strong enough differential sorting of types, the hazards of highly responsive types need to be sufficiently higher than the hazard rates of low-response types. The empirically observed differences in day-1 hazards between high- and low-response types that we have used for our simulations so far might simply be too small to yield decreasing population hazards as a result of differential sorting.

Figure XI therefore considers scenarios in which we increase the hazard rate of the high type. Panel A simply reprints the population hazard hazards with the types from the bottom part of Figure X. Panel B considers the case in which the high type has a day-1 hazard of 12%. Once more, we obtain increasing population hazards, independent of the high type's population share. Panel C depicts the population hazards for the case of a high-response type with a day-1 hazard rate of 16%. Even if 25% of the population would exhibit such a high hazard, we would again obtain increasing population hazards (solid grey line). Only if this high-response type would account for 50% or 75% of the population, one would eventually end up with weakly U-shaped or weakly decreasing population hazards (see the dashed grey and dotted black line in Panel C, respectively). However, the decrease is barely visible in the figure and less pronounced than the empirically observed decline in hazards during the first days of our D3 treatment. In addition, the implied day-1 population hazard levels (0.10 – 0.13) are well above the level observed in our data (0.05; see Figure 4 in the paper).[10]

A further dimension in which the simulations conflict with the data emerges when one uses the different types underlying Figure XI to simulate their behavior under a one-week the D1 treatment. Based on the types and population shares from Panel A and B above, the simulations yield increasing population hazards under a one-week deadline. For the most extreme type composition (Panel C), the resulting population hazards are essentially flat, with locally decreasing hazards only if the high type accounts for at least 50% of the population.

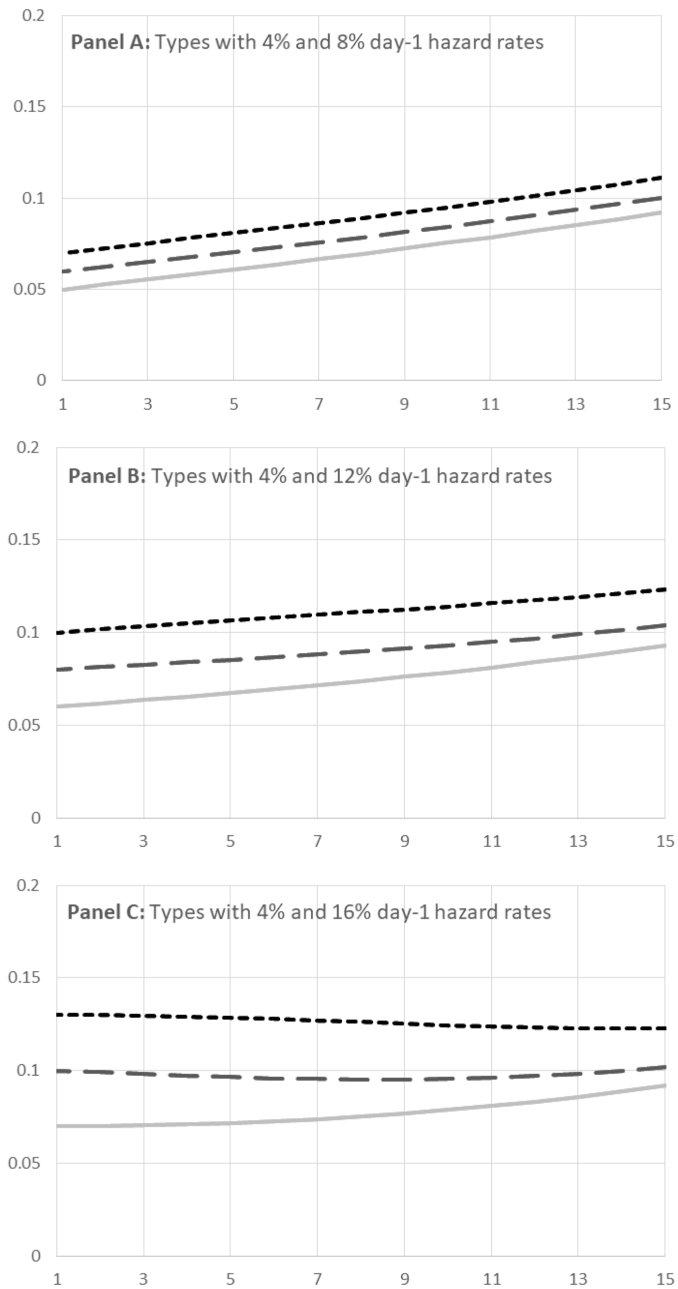
[9]More formally, let type g's costs be drawn from a distribution F_g that is an affine transformation of group j's distribution F_j (i.e., $F_g = a + bF_j$, with $a \in \mathbb{R}$ and $b \in \mathbb{R}^+$). As long as $\delta \approx 1$, any mixture of types g and j yields population hazards that are upward-sloping towards the deadline.

[10]To arrive at more pronounced local decreases in population hazards, one would need to increase the high type's initial hazards to even higher levels. Keep in mind, however, that (1) the day-1 hazards of the high type assumed in Panels B and C are already well beyond the maximum subgroup responsiveness that we observe empirically. In addition, such an exercise would yield day 1 population hazards that are even further away from the empirically observed level.

To arrive at the pronounced decline in population hazards that we empirically observe in the D1 treatment, one would need more extreme degrees of heterogeneity.

In a further step, we considered simulations with three types. In particular, we added a ‘very low’-response type which, by construct, helps to lower the level of population hazards. It turns out that the presence of this third type does not qualitatively alter the pattern from Panels A and B Figure XI. Only for the case from

Figure XI: Population hazard rates over 15 days (3-week deadline)



Notes: The figures illustrate population hazard rates (vertical axis) over 15 periods for mixtures of two types. Panels A, B, and C depict cases where the day-1 hazard rates of the high type are 2-, 3-, and 4-times the day-1 hazard rate of the low type (=4%), respectively. Within each panel, we consider mixtures of low/high-types with 75/25% (solid grey line), 50/50% (dark-grey, dashed line) and 25/75% (black, dotted line), respectively.

Panel C, in which the high type exhibits a day-1 hazard of 16% (and only if this high types are sufficiently prevalent), the decreasing or U-shaped population hazards already noted in Figure 3 become more pronounced. It is important to recall, however, that the latter scenarios are based on (i) a high type with an implausibly high day-1 hazard and a (ii) very large population share of the high type that, once more, (iii) implies a higher initial population hazard than what we observe empirically. Moreover, when we use these extreme three-type scenarios to simulate behavior under a one-week deadline, we observe in general again increasing population hazards.

Reexamination of Empirical Heterogeneity

To further explore empirically plausible boundaries of heterogeneity in our setting, we returned to our data and considered differences in initial responsiveness in fine-grained clusters of patients. In particular, we examined day-1 hazards in subgroups of patients that share the same combinations of up to 8 individual-level characteristics (including, among others, indicators for a patient's risk status, health insurance, 'history' at the dental clinic, spatial distance to the clinic). The results from this exercise corroborate that there simply is no sufficiently large group of very responsive types (that would be needed to generate decreasing or U-shaped population hazards) in our setting. For instance, if we split the sample in 16 different subgroups based on interactions of 4 individual-level characteristics, the highest observed day-1 hazard across all subgroups is 10%. This value is considerably below the initial responsiveness of 16% that we assumed for the simulations in Panels C of Figure XI above. Moreover, this most responsive subgroup accounts for less than 6% of the sample, which is an order of magnitude smaller than the share of high-response types that would be needed to generate decreasing or U-shaped population hazards in the numerical simulations. Considering even more fine-grained subgroup definitions – which, mechanically, also result in even smaller subgroups – or applying alternative approaches to measure patients' initial responsiveness (e.g., based on responses beyond the first day) does not alter these findings.

In sum, our simulation analysis indicates that it would require (a) strong differences in initial hazards across subgroups, and, additionally, (b) a large group of highly responsive types for differential sorting to result in U-shaped hazards under a 3-week deadline and decreasing hazards under a 1-week deadline. Both of these conditions need to be pushed well beyond what is empirically plausible in our setting.

C.4 Analysis of Cost Effectiveness

This section presents an illustrative analysis of the cost effectiveness of different deadlines and incentives in our main experiment. We first focus on the dentist's perspective and compare the treatments in terms of their costs to make a patient arrange a check-up appointment. For simplicity, we abstract from the patients' health-related benefits of check-ups and dental cleanings (see Section 2). We return to this point at the end of this section.

The different costs borne by the dentist are summarized in Table VII. They comprise the costs for printing, handling, and sending the reminder postcards (row 1) and the input costs for the different rewards in terms of material and personnel (rows 2 and 3).[11] While the former costs are borne independent of patients responding or not, the rewards only have to be paid out if patients book an appointment (in the ND treatment) and do so before the specified deadline (in the D1 and D3 treatment).

Table VII: Dentist's costs for reminders and rewards

Item	Cost in EUR
(1) Printing, handling, and sending postcards	0.47
(2) Small rewards	1.50
(3) Large rewards	40.00

Based on these costs and the patients' response rates (as depicted in Figure 2), we calculate the dentist's average costs per check-up appointment in the different treatments and over different time periods. Table VIII depicts the costs per check-up in the different treatment cell, when examining all responses within one week, three weeks, or 100 days after sending the postcard.

Our simple back-of-the-envelope calculation yields several interesting findings. First, it illustrates the value of deadlines, even if no reward is attached to them. Panel A of Table VIII compares costs in the no-reward condition between the one-week, three-week, and no-deadline treatment. When we focus on the responses within 21 days, the costs per arranged appointment are about 20–35% lower in the D1 and D3 treatment as compared to the ND treatment (EUR 2.14 in D1, 2.56 in D3, 3.11 in ND). Even after 100 days, the higher response rates in the two deadline treatments imply a 10–15% lower costs per arranged check-up than in the corresponding 'pure reminder' treatment that neither features a deadline nor rewards (EUR 1.41 in D1, 1.37 in D3, 1.59 in ND).

A second interesting insight from this analysis relates to an additional benefit from short deadlines: as documented in Section 4, they lead patients to respond earlier but do *not* lower the long-run response rates compared to the treatments with a longer or no deadline. Yet, the rewards only have to be paid for *timely* responses within the deadline. Panel C of Table VIII shows that this makes short deadlines very attractive from a cost-effectiveness perspective. While cumulative response rates after 100 days are rather similar in the D1 and the D3 treatment with large rewards (see upper panel of Figure 2), the dentist's costs per check-up in the D1 treatment with large incentive (EUR 23.65) are substantially below the costs in the D3 treatment (EUR 30.40). For both deadline treatments, in turn, costs are much lower than in the ND treatment (EUR 41.14), in which

[11] Note that the dentist's purchasing costs for the small rewards as well as his material and personnel costs for the dental cleanings are below the price faced by patients.

Table VIII: Costs per check-up

	7 Days	21 Days	100 Days
A. *No Reward*			
D1	3.42	2.14	1.41
D3	6.27	2.56	1.37
ND	5.94	3.11	1.59
B. *Small Reward*			
D1	6.23	3.98	1.85
D3	6.47	3.90	2.22
ND	9.57	5.42	2.86
C. *Large Reward*			
D1	41.75	35.60	23.65
D3	42.25	41.29	30.40
ND	44.20	42.27	41.14

Notes: The table presents the dentist's average costs per check-up appointment, based on empirically observed response rates in each treatment cell (Figure 2), the costs for sending postcards to all patients, and disbursed rewards for patients who respond within the deadline.

all patients who eventually respond are eligible for receiving the (costly) reward. This shows that appropriately chosen deadlines may have three benefits: (1) they can induce earlier responses (2) without lowering overall response rates, and (3) they do so at lower overall costs than comparable incentives which are not time-limited.

Finally, holding deadlines constant and focusing on the incentive dimensions, the numbers from Table VIII illustrate that offering a large reward (the free dental cleaning) only pays off if the non-pecuniary benefits of patients responding early and at an overall higher rate are substantial. While the 100-day response rates in the large-reward condition are roughly 15 percentage points higher than in the conditions with no or low rewards (cp. Figure 2), the costs per check-up appointment are between 20-40 EUR higher, depending on the length of the deadline.

To gauge whether the non-pecuniary benefits from additional check-ups exceeds this threshold, it is useful to consider which benefits are omitted in our analysis. First, the higher number of check-ups in the large-reward treatment leads to health benefits for the additional patients responding to the intervention. The same holds for the higher rate of professional dental cleanings, which themselves have positive health effects for patients (as well as potential indirect economic returns related to beauty; cp. Glied and Neidell 2010). Second, in the treatment with large rewards and deadlines of 1 or 3 weeks, we observe a particularly high number of patients responding earlier than in the ND condition (cp. Figure 2). These early responses are likely to yield additional health benefits as many dental diseases are easier and cheaper to treat if discovered at an early stage. Finally, the large-reward intervention can have long-term benefits for the dentist (and patients) due to an increase in 'customer loyalty'. Adding price tags to these different benefits is beyond the scope of our analysis. It appears conceivable, however, that the (expected) benefits may add up to a total of more than EUR 20. This suggests that the relatively costly high-reward condition could be effective from a cost-benefit perspective, at least when coupled with a relatively tight deadline.

C.5 Online Survey

We conducted a large online survey experiment on individuals' perceptions of deadlines (N = 3,078) and a smaller post-experimental survey at the dental clinic (N = 273). For the latter survey (which was administered after the follow-up experiment), we invited patients in the waiting room of the clinic to participate in a short (approx. 5 min) structured face-to-face interview. Questions covered individuals' views on dental health prevention, appointment planning, and the perception of deadlines and check-up reminders. The online survey experiment was conducted in collaboration with a professional survey provider that maintains a sample that is representative for Germany's adult population. After a number of introductory questions, participants in the online survey were assigned to a vignette scenario in which they were shown one of the postcards from our experiment ('Imagine you receive the following postcard from your dentist...'). Across survey participants, we randomly varied the postcard texts, such that they corresponded to one of the treatment cells from the main or follow-up experiment. The survey then asked for individuals' perceptions of the respective postcard along various dimensions. Tables IX and X summarize participant characteristics and the main outcomes of the vignette experiment.

Table IX: Summary Statistics – Online Survey

	D1	D3	D6	D10	ND	Total
Age	37.77	39.36	37.65	38.09	38.58	38.47
	(13.32)	(14.05)	(12.95)	(13.37)	(13.13)	(13.46)
Female	0.55	0.57	0.64	0.54	0.56	0.56
	(0.50)	(0.50)	(0.48)	(0.50)	(0.50)	(0.50)
PrivateHI	0.20	0.20	0.20	0.20	0.19	0.20
	(0.40)	(0.40)	(0.40)	(0.40)	(0.39)	(0.40)
N	879	868	213	225	893	3078

Notes: The table presents summary statistics (means and standard deviations in parenthesis) for participants in the online survey, as well as the outcome from the random assignment of survey participants to vignette scenarios with different deadline lengths. Note that within the vignettes for treatments D1, D3 and ND, we also (randomly) varied the reward level. Among participants who face the D1 treatment (and analogously for D3 and ND), respondents are equally likely to encounter a large-, a small-, or no-rewards (as in our main experiment) or an intermediate reward level (as it is used in the follow-up experiment). For treatments D6 and D10, we only used the intermediate reward level. Hence, these two treatment cells are only populated by a 1/4 of the observations in each of the three treatments that was used in both the main and follow-up experiment. To resemble our experimental population in terms of key observables, the survey used sampling quota for gender (56.2% females), private health insurance status (19.5%) and age (31% with age $<$ 30 years; 30% with 30 $\leq$ age $<$ 40; 18% with 40 $\leq$ age $<$ 50; 11% with 50 $\leq$ age $<$ 60; and 10% with age $\geq$ 60).

Table X: Online Survey – Results of Vignette Study

	D1	D3	D6	D10	ND	Total
Based on the specific content of the postcard...						
(1) *Would you draw any conclusions about the dentist's competence?*						
Yes - positive	28.33	27.49	26.42	26.67	26.69	27.36
No	65.48	65.89	66.04	65.33	67.23	66.13
Yes - negative	6.19	6.61	7.55	8.00	6.08	6.51
(2) *My dentist wants to increase showup / advertise the clinic*						
(Totally) Agree	56.28	55.04	56.13	57.78	55.39	55.77
(Totally) Disagree	43.72	44.96	43.87	42.22	44.61	44.23
(3) *My dentist wants me not to postpone or forget arranging a new check-up*						
(Totally) Agree	90.87	90.38	93.87	91.11	91.46	91.13
(Totally) Disagree	9.13	9.62	6.13	8.89	8.54	8.87
(4) *My dentist thinks that I have an acute dental health problem*						
(Totally) Agree	11.87	11.36	7.55	7.11	12.47	11.25
(Totally) Disagree	88.13	88.64	92.45	92.89	87.53	88.75
What do you think about the deadline?						
(5) *The deadline puts me under pressure when arranging a new appointment*						
(Totally) Agree	39.61	33.96	32.08	25.24	—	35.16
(Totally) Disagree	60.39	66.04	67.92	74.55	—	64.86

Notes: The table tabulates response for vignette scenarios with different deadline lengths. For ease of exposition, we summarized responses to questions (2)–(5) from a four-point scale (totally agree / rather agree / rather disagree / totally disagree) to a binary scale. Sample size is $N = 3,066$ (for questions 1–4) and $N = 2,159$ (for question 5, which was not asked in the ND treatment). Due to the complementary variation in reward levels (see note for Table IX), response rates can only be compared between D1, D3, and ND as well as between D6 and D10. For a comparison between the three treatments that were included in both main and follow-up experiment and the two treatments that were solely used in the follow-up experiment, one has to control for the reward level.

Table XI: Online Survey – Further Results

(1) *How often do you intend to have dental check-ups?*	
$>$ 2 times per year	10.40
2 times per year	46.04
once a year	35.09
$<$ once a year	4.61
never (only in case of pain)	3.87
(2) *Number of dental check-ups during the past 2 years*	
0	6.56
1	10.77
2	31.19
3	17.77
4	26.25
$>$4	7.46
(3) *How pleasant / unpleasant do you consider professional dental cleanings?*	
very pleasant	5.14
rather pleasant	15.29
neither / nor	43.98
rather unpleasant	30.35
very unpleasant	5.24
(4) *Would you make use of a professional dental cleaning if it was offered free of charge by your health insurance plan?*	
definitely yes	67.26
rather yes	25.70
rather not	5.80
definitely not	1.23
(5) *What do you think is the value of the 'small present'?*	
$\sim$ EUR 2.50	72.90
$\sim$ EUR 5	21.96
$\sim$ EUR 7.50	1.40
$\sim$ EUR 10	2.34
$>$ EUR 10	1.40

Notes: The table tabulates responses (in percent) to further questions from the online survey. Sample size ranges from $N = 3,078$ to $N = 3,017$ (minor attrition). Question (5) was administered only to participants in the vignette scenario with small rewards ($N = 642$).