Department of Economics

University of Reading

Managing Excess Demand for Primary Care: Evidence from Online Experiments

by Diya Abraham, Ondrej Krcal

and Jonathan Stäbler

Discussion Paper No. 2026-04

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6AA
United Kingdom

www.reading.ac.uk

Managing Excess Demand for Primary Care: Evidence from Online Experiments

Diya Abraham[*], Ondřej Krčál[†], Jonathan Stäbler[‡]

May 26, 2026

Abstract

Primary healthcare systems in many developed countries are under strain, partly due to unrestricted patient demand. In response, policymakers have introduced measures to curb unnecessary GP visits, including (i) instituting a small upfront fee for GP visits, (ii) implementing a self-report based triage system, and (iii) providing more information to patients about their condition before they make an appointment with their GP. We evaluate the effectiveness of these approaches using two online experiments with a representative sample of UK adults. The first experiment involves induced monetary incentives in a laboratory-style study while the second is a health-framed vignette study. We find that while all three interventions are effective in the laboratory study, only the intervention that provides patients with more information about their condition reduces low-priority demand in the vignette study. We discuss implications for policy and for the study of health-related decision-making.

Keywords: Health Care Systems; Common Pool Dilemma; Type Uncertainty; Online Experiment

JEL Classification: I1, D8, D9, C9

[*]University of Reading, Shinfield Rd, Reading RG6 6EL, United Kingdom, e-mail: d.e.abraham@reading.ac.uk

[†]Masaryk University, Lipova 41A, 60200 Brno, Czech Republic, e-mail: ondrej.krcal@econ.muni.cz

[‡]Masaryk University, Lipova 41A, 60200 Brno, Czech Republic, e-mail: jonathan.stabler@econ.muni.cz

I Introduction

In many high-income countries with fully tax-financed health systems, the demand for primary healthcare services has outpaced supply. General Practitioners (GPs) are typically a patient's first point of contact, yet there is a serious shortage of GPs in many of these countries, particularly in the UK.[1] Declining GP-patient ratios are associated with longer waiting times and lower quality and continuity of care, ultimately leading to poorer health outcomes (Swami et al., 2018; Freedman et al., 2021; Vallejo-Torres and Morris, 2018; Shurtz et al., 2022; Kajaria-Montag and Freeman, 2020; Kajaria-Montag et al., 2024).[2] One potential way to solve this problem would be to increase GP supply by training and retaining more GPs, however this is difficult to achieve in the short-term (Batavanis et al., 2025).

In a health system where access is free at the point of service, and where, as a result, there is considerably higher utilization of primary care services (Iizuka and Shigeoka, 2022; Aron-Dine et al., 2013), the problem might also be addressed by reducing unnecessary demand stemming from moral hazard, mistakes and need-uncertainty (Finkelstein et al., 2019; Zweifel and Manning, 2000; Baicker et al., 2015; Dardanoni and Wagstaff, 1990). Specifically, healthcare systems may alleviate the pressure on care by introducing incentives for patients to engage in a form of self-triage, thereby reducing the number of low-need patients that enter the system in the first place. In this paper, we use two experiments with a large representative sample of UK adults to examine the causal effect of three common interventions to better manage demand: (i) a small fixed payment for GP visits, (ii) a self-report based triage system that introduces a moral cost of misreporting symptoms, and (iii) improving patients' knowledge of their health condition.

Many countries with universal health coverage have rolled out equivalent policy measures with the aim of better managing the demand for primary care. Introducing or increasing co-payments for GP visits in particular is quite common.[3] Empirical work in this area typically

[1] In England, the number of patients registered with a GP rose by 10% over the past decade, while the number of GPs per 100,000 patients fell by 15% (Care Quality Commission, 2024; The Nuffield Trust, 2025). Reasons for lower GP access in the UK include chronic underinvestment, GP burnout, falling job satisfaction, high workload, fewer GPs choosing to work full-time, and warped incentives for GP surgeries under the Additional Role Reimbursement Scheme (ARRS) (Fisher, 2026).

[2] On average across ten OECD countries, 18% of survey respondents reported waiting more than a week to see a GP or a nurse, and 57% waited two days or longer (OECD, 2025). When treatment is delayed, patients' productivity and quality of life is reduced. Limited access to GPs can also lead individuals to seek out care in emergency care departments, resulting in higher system costs and worse outcomes for those with more urgent medical needs (Blunt et al., 2015; Cowling et al., 2014; Dolton and Pathania, 2016; Hoe, 2022).

[3] Countries with universal health coverage that have implemented small co-payments for GP visits include Ireland, Finland, Norway, Sweden, France, Portugal, Latvia and Slovenia. Countries that have trialed and subsequently abolished co-payments include Germany (Bauer et al., 2006; Schreyögg and Grabka, 2010), Israel (Rosen et al., 2011), Hungary, Czech Republic and Slovakia (Baji et al., 2010). In England, patients incur a fixed fee per prescription rather than per GP visit, which is also a fairly common practice in many European

exploits policy changes using administrative health data (see for example Jakobsson and Svensson, 2016; Iizuka and Shigeoka, 2022). However, identifying the effectiveness of such policies is challenging. This is because individual health needs are heterogeneous and imperfectly observed, individuals who forgo care altogether are not captured in these datasets, and policy changes themselves are endogenous to the state of the health system and shaped by the institutional and behavioral context in which they are implemented. The current study addresses this gap by using two complementary experiments with random assignment to different policy regimes, which allows us to identify causal effects on low-priority demand for limited health resources. In both experiments, we elicit the choices of participants with heterogeneous but uncertain needs for health services, allowing us to observe both utilization and non-utilization. In the first experiment, these choices are realized as extraction decisions in a common-pool resource game with induced monetary incentives, while in the second, they are elicited as stated care-seeking decisions in a framed vignette experiment.

In the laboratory experiment, we model the consequences of over-utilization in a supply-constrained health system within the common-pool resource framework. We limit the number of feasible extractions from the common pool to $n-1$ in groups of size n. This means that if all members of a group choose to extract resources (i.e., make an appointment with their GP), one individual is randomly prevented from extracting (i.e., denied timely care). To capture heterogeneity in medical need, participants are randomly assigned a health type – high-harm or no-harm – where only high-harm individuals incur a loss if they do not (or are unable to) extract. Prior to making their choice, participants do not know their health type but receive a private noisy signal about it, which represents the symptoms they might experience before deciding whether to visit their GP.[4]

To examine the impact of existing policy measures on the decision to extract in this context, we collect data from a *Baseline* (described above) and three treatments: In the *Cost* treatment, there is an upfront fee for extraction similar to a co-pay; in the *Triage* treatment, participants report being high-harm or no-harm after receiving their signal, and are aware that reporting being high-harm is equivalent to choosing to extract; in the *Accuracy* treatment, participants receive a more accurate signal about their health type relative to the other treatments. Across all treatments, the consequences of extracting or not extracting (equivalently, reporting being high-harm or no-harm) is the same and depends on one's true type.

In Experiment 2, participants view five hypothetical health scenarios and report how likely

countries.

[4]By introducing type-uncertainty and private signals about players' types prior to their extraction choice, we extend the Common Pool Resource (CPR) game (Ostrom et al., 1994) to model extraction decisions in an overwhelmed healthcare system – or, more broadly, any resource dilemma where moral hazard coexists with uncertainty about one's benefits from consuming the resource.

they would be to make an appointment with their GP in each one. Four of the five scenarios outline common symptoms that were assessed by a sample of fourteen certified GPs in the UK as being very unlikely to require a GP visit. We refer to these as the no-harm scenarios. The fifth scenario describes a health condition that the aforementioned GPs assessed as being highly likely to require a GP visit, i.e., a high-harm scenario. Prior to making their choices, all participants are informed that while GP visits are tax-financed in their country and thus free at the point of service, the health system is overwhelmed and as a result many individuals with serious conditions find it difficult to access timely care. As in the laboratory study, participants are randomly assigned to different policy regimes in which they (i) incur a small fee when visiting their GP (*Cost* treatment), (ii) need to exaggerate their symptoms in order to secure a GP appointment (*Triage* treatment) and (iii) both of the above (*CostTriage* treatment). To create the *Accuracy* treatment in this study, we varied the amount of relevant information participants received about their health condition in each scenario. The additional details shown in the *Accuracy* condition were designed to resemble information individuals might obtain through self-reflection, which might for example be triggered through interaction with a triage nurse, AI chatbot or a friend with a background in medicine.

We find that in Experiment 1, when the system is overwhelmed (i.e., when participants are aware that all others in their group have chosen to extract), all our treatments reduce the extraction rate of no-harm signal receivers. The *Accuracy* treatment has a smaller effect compared to the *Triage* and *Cost* treatments, which reduce the extraction rate to the same extent. In Experiment 2, however, we find that only the *Accuracy* treatment, in which subjects receive more information about their symptoms, reduces participants' self-reported likelihood to make a GP appointment in the low-harm scenarios. The *Cost*, *Triage* and *CostTriage* treatments have no impact.

While the health system is made more efficient when no-harm signal receivers choose *not* to extract, the opposite is true in the case of the high-harm signal receivers. In both studies, we thus also explore whether our treatments *increase* participants' extraction choices after receiving a high-harm signal. We find that in Experiment 1, high-harm signal receivers are indeed more likely to extract in the *Triage* treatment compared to both the *Baseline* and *Cost* treatments. This is consistent with previous work that cost-sharing deters appropriate care and speaks to the benefit of instituting a triage system over a co-pay. The *Accuracy* treatment, meanwhile, does not increase extraction by the high-harm signal receivers, but results in the highest efficiency overall by reducing extraction errors. Turning to Experiment 2, we find that in the high-harm scenario, more information about one's symptoms (i.e., greater accuracy) increases the reported likelihood of making a GP appointment; but that the *Cost* and *Triage* treatments have no positive impact.

Thus across both studies, there is evidence that the highest efficiency is achieved when patients receive more information about their condition. In the laboratory study, this happens both by reduced extraction following no-harm signals and because higher signal accuracy reduces errors among individuals receiving no-harm signals while in the vignette study it is the direct result of better patient decision-making across both the high and low-harm scenarios.

The first contribution of this paper is to research that examines how to reduce moral hazard in the demand for primary care through the use of co-payments. The evidence so far is mixed. While most studies find that co-pays reduce primary care utilization, it is unclear whether this reduction comes from those who over-use or under-use primary care services, and whether it results in better or worse health outcomes overall (see Kiil and Houlberg (2014) for a review). Iizuka and Shigeoka (2022) exploit the rapid expansion of child subsidies in Japan to show that while small co-payments reduce primary care utilization among healthier children, they also decrease screening for depression and ADHD, which are cost-effective preventive interventions for this age group. Under-utilization in response to increasing *prescription* co-pays has also been documented by Chandra et al. (2024), and Goldman et al. (2006) who show that increasing co-payments for heart disease and cholesterol-lowering medication have strong adverse effects on precisely those individuals who are at high risk for these diseases.

These and other studies investigating the effect of introducing or increasing co-pays in primary care employ quasi-experimental data from policy changes at the national level. Since such policy changes are endogenous and shaped by the state of the health system, their effects depend on the institutional setting, the health status of the population, prevailing care-seeking behavior, and any concurrent reforms. While the authors use credible difference-in-difference (Cockx and Brasseur, 2003; Chiappori et al., 1998) or regression discontinuity designs (Landsem and Magnussen, 2018; Layte et al., 2009), the fact remains that they cannot *directly* observe how individuals with heterogeneous needs for primary health care respond to the policy change. We address this limitation through an experimental design that induces heterogeneous need and noisy signals, while allowing us to observe both extraction and non-extraction choices under randomly assigned policy regimes. We find that while a co-pay reduces low-priority demand when decision making is in the domain of money, it does not impact behavior when decision making is about health. The latter result is consistent with research that finds there is limited, if any, long-term impact of introducing a co-pay on the frequency of GP visits (Schreyögg and Grabka, 2010; Chiappori et al., 1998; Jakobsson and Svensson, 2016) or on overall efficiency (Cockx and Brasseur, 2003).

Our second contribution concerns the use of triage systems as a tool to manage demand by creating the need for patients with mild symptoms to lie or exaggerate the severity of their symptoms in order to be seen. This feature of a triage system is relatively unexplored in the

health literature since the underlying assumption in most empirical and theoretical work on triage systems in healthcare is that patients will always be truthful in reporting their symptoms (Guan et al., 2025; Çakıcı and Mills, 2021). However, if there is some level of moral hazard in primary care utilization, we can assume that some patients who visit their GP are aware that doing so is likely to be deemed unnecessary on average. With this in mind, along with the experimental economics literature on individuals' inherent preferences for truth-telling (see e.g. Abeler et al. (2014); Gneezy et al. (2013); Thami (2025) and a recent meta-analysis of experimental studies on truth-telling by Abeler et al. (2019)) we conjecture that simply asking patients to truthfully report their symptoms (equivalently, their signal) instead of asking them to directly report whether they would make a GP appointment (equivalently, to extract) is sufficient by itself to reduce the likelihood of unnecessary extractions.[5] Thus, rather than quantifying potential efficiency gains from a more accurate patient ordering by need, the *Triage* treatments in both our experiments are specifically designed to study the impact of having to exaggerate/mis-report one's symptoms in order to make a GP appointment.

We find that while our *Triage* treatment is effective at reducing unnecessary extraction in Experiment 1, both triage-based treatments, i.e., *Triage* and *CostTriage*, have no impact on the likelihood of making a GP appointment in Experiment 2. This lack of an effect in the domain of health-related decision-making is in line with anecdotal evidence indicating that people are indeed willing to exaggerate their symptoms in order to see their GP sooner (Heather, 2017) or secure an ambulance faster (Jones, 2020). The effect of the *Triage* treatment in the laboratory study, meanwhile, is consistent with previous laboratory evidence that individuals exhibit an intrinsic preference for truth-telling when decisions are about money.

Our *Accuracy* treatment speaks to the use of *informative* triage systems and AI-powered software that provide patients with more information about their health needs. Most GP practices in the UK and Ireland already employ such informative triage systems wherein patients fill in an online form that gives them more information about their health condition while simultaneously helping to screen patients through an implicit honor system. However, our results suggest that it may be useful to separate these two functions. To this end, an AI-based symptom assessment app that is not tied to a GP surgery's appointment scheduling system might be more effective. "Ada" (by Ada Health) is one such widely-used app that has been received positively by patients in a primary care setting (Miller et al., 2020). Many others exist. The UK's NHS is currently developing its own AI-powered triage and guidance within its official NHS App that will help patients decide when to seek GP care (Global Government Forum, 2026). We believe it might be useful to have this in-app AI tool be operationally distinct

[5] In an experiment where customers can make strategic claims to reduce their waiting times, Estrada Rodriguez et al. (2025) finds that there are indeed certain conditions under which an "honor-system", where service priority is given according to customer claims, would work.

from the appointment scheduling system.

This paper also contributes to the literature employing laboratory experiments to examine decision-making in primary and secondary healthcare settings. Most work in this area relies on incentivized experiments to study patients' health insurance choices (Kairies-Schwarz et al., 2017; Hermanns et al., 2025) or supply-side issues, such as physicians' incentives (Brosig-Koch et al., 2017, 2024). In contrast, we focus on policy interventions aimed at better managing patient demand, complementing our incentivized laboratory experiment with a health-framed vignette study.

The rest of this article is organized as follows: Section II details the theoretical considerations, Section III presents the experiment and results of Experiment 1, Section IV presents Experiment 2, Section V provides a discussion, and Section VI concludes.

II Theoretical Considerations

In this section, we introduce our theoretical model and derive testable predictions. We adapt a one-shot common-pool resource game to the problem of limited supply of primary healthcare. Let n be the number of group members, who can be high-harm types (θ_h) with probability $h > 0$ or no-harm types (θ_{no}) with probability $1 - h$. The probabilities and the harm types are common knowledge but the individual realizations are unknown to the players. Instead, each player receives a private noisy signal $\sigma \in \{\sigma_h, \sigma_{no}\}$ about their own harm, where signal precision is given by $k \in (0.5, 1)$. Given their signal, σ, players use Bayesian updating to determine their posterior probability $p_h(h, \sigma, k)$ of being a high-harm type.[6]

All players simultaneously decide whether to extract resources ($x_i = 1$) or not ($x_i = 0$) from the common pool. While no-harm types do not face any losses during the game, high-harm types are only able to *avoid* a loss of A if they extract resources from the common pool.[7] However, as the common pool is limited, only $n - 1$ players can extract resources from it at any given point. If all n group members decide to extract, one is randomly excluded, with $z_i \in \{0, 1\}$ indicating whether player i ultimately extracts. The resulting payoff for each group member is given by:

$$\pi_i := \begin{cases} 0 & \text{if } z_i = 1 \\ -p_h(h, \sigma, k)A & \text{otherwise,} \end{cases}$$

The timing of the game is as follows. First, each player's type, high-harm or no-harm, is determined. Next, all players receive private signals about their harm-type. Then, they decide

[6] $p_{\text{high}}(\theta_h|\sigma_h) = \frac{P(\sigma_h|\theta_h)P(\theta_h)}{P(\sigma_h|\theta_h)P(\theta_h)+P(\sigma_h|\theta_{no})P(\theta_{no})} = \frac{kh}{kh+(1-k)(1-h)}$ and $p_{\text{high}}(\theta_h|\sigma_{no}) = \frac{(1-k)h}{(1-k)h+k(1-h)}$

[7] Note that though we normalize the loss incurred by no-harm types to 0, the theoretical results readily extend to no-harm types also incurring a loss of $B < A$.

simultaneously whether to extract resources from the common pool. If all n members of a group decide to extract, one of them is randomly prevented from extracting. If less than n members of a group decide to extract, all those who chose to extract, do so.

We focus on individual extraction choices in the situation where all other members of one's group have chosen to extract. In this way, we create an overcrowded system where own extraction choices necessarily lead to the exclusion of one member of the group.[8] In this situation, the socially optimal solution requires that no-harm players do not extract so that high-harm players can avoid the loss of A. But since players face uncertainty about their type, the best social solution is achieved when only those experiencing severe symptoms (i.e., receiving a high-harm signal) choose to extract.[9]

To study the optimal extraction choices from a given individual's perspective, we employ risk-neutral agents with altruistic preferences (Andreoni and Miller, 2002). Such agents experience *dis*utility if their extraction leads to the exclusion of a high-harm type in their group. For simplicity, we assume that players derive disutilty only from the one excluded player (denoted by j) rather than from the payoffs of all group members. Let $\alpha_i \in [0, 1)$ denote how much players care about the loss of the excluded player. α can also be interpreted as a preference for norm adherence, where the expected cost of deviating is represented by the expected externality. In a Triage system – where players can only extract if their self-reported symptoms are high enough – we assume that players have intrinsic preferences for truth-telling (Abeler et al., 2019) and therefore also experience a psychic cost of exaggerating their symptoms, i.e., reporting being high-harm after receiving a no-harm signal. Finally, players may incur a cost c_i when extracting, which represent either a monetary co-payment or the non-monetary burdens of seeking care (e.g., time costs, commuting costs, or infection risks). We thus have the following utility specification:

$$U_i = \pi_i + \alpha_i \pi_j \mathbb{1}_{z_i=1} - c_i \mathbb{1}_{z_i=1} - c_i^{psych} \mathbb{1}_{Triage \wedge \sigma_{no}} \tag{1}$$

Where π_i represents player i's own payoff, π_j the payoff of an excluded other player j if player i gets to extract (i.e., if $z_i = 1$), α_i the individual level of altruism, c_i the (monetary and non-monetary) extraction costs if player i gets to extract (i.e., if $z_i = 1$), and c_i^{psych} the individual psychic costs of reporting being high-harm despite receiving a no-harm signal in a Triage system. Player i chooses x_i in order to maximize expected utility:

[8] Rather than examining equilibrium selection between symmetric (all extract) and asymmetric (all but one extracts) outcomes, we study individual choices in this situation where all other members of one's group have chosen to extract. This further eliminates the potential coordination problem that would have otherwise come up in this setting.

[9] This is true as long as the extraction fee c is lower than the expected benefit of extracting, which we assume is always true for high-harm players.

$$\arg\max_{x_i \in \{0,1\}} \pi_i + \alpha_i \pi_j \mathbb{1}_{z_i=1} - c_i \mathbb{1}_{z_i=1} - c_i^{psych} \mathbb{1}_{Triage \wedge \sigma_{no}}$$

The following proposition characterizes the necessary and sufficient condition for player i to extract.

Proposition 1 (Optimal extraction rule)**.** *Suppose a risk neutral and altruistic player i faces an extraction decision in the scenario as described above. Then player i extracts iff:*

$$\underbrace{\frac{n-1}{n} p_h(h, \sigma, k) A}_{\text{Benefits of decision to extract}} > \underbrace{\alpha_i \frac{n-1}{n} hA}_{\text{Altruistic cost}} + \underbrace{\frac{n-1}{n} c_i}_{\text{Extraction cost}} + \underbrace{c_i^{psych} \mathbb{1}_{Triage \wedge \sigma_{no}}}_{\text{Lying cost}} \tag{2}$$

Proof. See Appendix A.A. □

Altruistic players deciding whether to extract trade off the benefit of avoiding the loss of A (if they end up being high-harm) against the cost of excluding a high-harm player, alongside any extraction and psychic costs. The benefits of extracting depend on whether player i is selected to extract, which happens with probability $\frac{n-1}{n}$. The costs include harming a potentially high-harm player, who could be excluded from extracting with probability $\frac{n-1}{n}$. Since player i does not observe others' signals, the probability that the excluded player is high-harm is the prior probability h.

II.A No-harm Signal

We now examine the comparative statics of three proposed policy measures to reduce low-priority demand (i.e., extraction after receiving a no-harm signal $\sigma = \sigma_{no}$) in an overwhelmed system (i.e., when all others in one's group have chosen to extract).

The first corollary describes how extraction choices respond to an increase in the signal accuracy, k. A more accurate signal makes it more likely that no-harm signal receivers are actually no-harm types. Increased signal accuracy thus decreases their expected benefit from extraction. Consequently, players who would extract under low signal accuracy may switch to non-extraction when signal accuracy increases.

Corollary 1 (Effect of signal accuracy)**.** *An increase in signal accuracy, k, weakly decreases the frequency of extraction after receiving a no-harm signal.*

Proof. See Appendix A.B. □

The next corollary describes the impact of introducing a co-payment, or more generally an increase in extraction costs c_i. Since such an increase leaves the benefits of extraction unchanged, it reduces the net-benefits of extracting.

Corollary 2 (Effect of extraction costs)**.** *An increase in the extraction costs c_i weakly decreases the frequency of extraction after receiving a no-harm signal.*

Proof. See Appendix A.C. □

Corollary 3 describes the effect of a self-report based triage system, where only patients reporting they they are high-harm types are considered for extraction. This creates an incentive to always report being high-harm. However, doing so may create psychic moral costs from over-reporting when receiving a conflicting (no-harm) signal.

Corollary 3 (Effect of moral misreporting costs)**.** *An increase in the psychic costs c_i^{psych} of reporting being high-harm after receiving a no-harm signal weakly decreases the frequency of extraction.*

Proof. See Appendix A.D. □

II.B High-harm Signal

We now study extraction choices after receiving a high-harm signal. Since players cannot over-report their harm type in this scenario, the psychic cost of mis-reporting does not exist, allowing us to set $c_i^{psych} = 0$. Further, receiving a high-harm signal increases the probability that they are actually a high-harm type – also relative to other members of their group. Consequently, individuals will always extract as long as (i) they care more about their own expected payoff compared to that of another member of their group (i.e., if $\alpha \leq 1$) and (ii) extraction costs c_i are sufficiently low. The following corollary formalizes these conditions.

Corollary 4 (Extraction after a high-harm signal)**.** *For small enough extraction costs $c_i < (p_h(\theta_h|\sigma_h) - \alpha_i h)A$, Player i always decides to extract after receiving a high-harm signal.*

Proof. See Appendix A.E. □

III Experiment 1: Laboratory Experiment

Experiment 1 is a laboratory-style experiment that uses induced monetary incentives to study the impact of three different policy regimes on the demand for primary healthcare services. It directly implements the theory presented in Section II and tests the effects of the policies on individual extraction choices in an overwhelmed system.

III.A Experimental Conditions and Predictions

In Experiment 1, we set the potential loss to $A = 2.5$ and signal accuracy to $k = \frac{2}{3}$. We use four between-subject treatments to test the proposed healthcare system measures (see Table 1). In the Baseline, players can decide to extract without the need to over-report their type or pay a fee when extracting. Thus, we set $c_i = 0$ and $c_i^{psych} = 0$. In the *Accuracy* treatment, we change the precision of the signal to $k = \frac{9}{10}$. In the *Cost* treatment, we introduce an extraction fee and set $c = 0.5$. Finally, in the *Triage* treatment, players have to report being high-harm in order to be able to extract. Thus, we introduce the psychic costs of needing to exaggerate one's symptoms (i.e., report being high-harm) after receiving a no-harm signal, i.e. $c_i^{psych} > 0$.

TABLE 1: Experiment 1 – Overview of pre-registered predictions after receiving a no-harm signal.

	Prediction	Derivation
Baseline (k = $\frac{2}{3}$, $c_i = 0$, $c_i^{psych} = 0$)		
	No specific prediction	
Accuracy Treatment (k = $\frac{9}{10}$, $c_i = 0$, $c_i^{psych} = 0$)		
Prediction 1	*Extraction choices decrease with higher signal accuracy after receiving a no-harm signal compared to the Baseline.*	Corollary 1
Cost Treatment (k = $\frac{2}{3}$, $c_i = 0.5$, $c_i^{psych} = 0$)		
Prediction 2	*Extraction choices decrease with the introduction of $c = 0.5$ after receiving a no-harm signal compared to the Baseline.*	Corollary 2
Triage Treatment (k = $\frac{2}{3}$, $c_i = 0$, $c_i^{psych} > 0$)		
Prediction 3	*Extraction choices decrease with the introduction of $c_i^{psych} > 0$ after receiving a no-harm signal compared to the Baseline.*	Corollary 3

Table 1 summarizes our pre-registered predictions for each of our treatments. The main takeaway is that we expect a decrease in extraction decisions in all treatments compared to the *Baseline*. As the strength of this decrease depends on the distribution of α, the size of c_i^{psych}, as well as risk preferences, we remain ex-ante agnostic about the quantitative and relative decrease across the treatments.

III.B Experimental Procedures

The experiment was programmed in oTree (Chen et al., 2016) and conducted online on Prolific.[10] Participants first read the experimental instructions about the common pool resource paradigm (see Appendix B.F). The instructions contained information about the size of their randomly assigned groups (four members each) as well as the chance that they were a no-harm or high-harm type (50%-50%), and detailed the associated payoffs from extracting from the common

[10]The experiment was pre-registered at https://aspredicted.org/q5mc-9k6y.pdf and received approval from the Research Ethics Committee at the University of Reading.

pool in each case (i.e., a loss of £2.50 if they were high-harm but ended up not extracting, and no loss if they were no-harm). They were also told they would receive a private (noisy) signal about their type, and that if all four members of their group chose to extract, one would be randomly excluded from extraction.

In the *Cost* treatment, they learned about the extraction fee (i.e., £0.50) they would incur regardless of their actual type should they end up extracting. In the *Triage* treatment, rather than deciding whether or not to extract, they were informed they would be asked to report whether they were high-harm or no-harm after receiving their signal with only those reporting they were high-harm being considered for extraction. In the *Accuracy* treatment, they were made aware that there was a 90% chance that the private signal about their type was correct. In the other three treatments, they knew that there was just a two-third (67%) chance the signal was correct.

After reading the instructions, participants answered 18 comprehension questions.[11] If the answer a participant provided was incorrect, they were provided with a short explanation and then asked to try again. They could not proceed until they had answered all questions correctly. The main goal of these questions was to improve participants' understanding of the instructions without discouraging continued participation. We recorded whether the response provided was correct in the first instance, and use this information in our robustness analysis.

Participants' extraction decisions (equivalently, the decision whether or not to report they were high-harm in the *Triage* condition) were elicited via the strategy method. Each participant made a total of six decisions. Of these, two were unconditional extraction (reporting) decisions, in which participants chose whether or not to extract (report they were high-harm) without any information about the extraction (reporting) decisions of the other three members of their group, for each of the two possible private signals they could receive about their type. Each of the other four decisions were conditional decisions depending on whether (i) all the three other members of the group or less than three other members of their group had decided to extract (equivalently, report that they were high-harm in the *Triage* condition) and (ii) they themselves had received a no-harm or a high-harm signal.

Figure 1 shows the first of the four conditional decisions in the *Baseline*. Our pre-registered hypotheses are based on just this first decision (i.e., no-harm signal and all other group members extracting). The remaining five decisions (three conditional and two unconditional) were elicited to ensure that they were incentivized to provide their true choice in that first decision.[12] We

[11] In the *Cost* treatment, there was one additional comprehension question about the extraction fee.

[12] After the experiment, participants' earnings in each treatment were calculated in four steps: (1) first, within each treatment, all participants were randomly assigned to groups of four; (2) next, their actual harm type (high-harm or no-harm) was drawn with each type being equally likely, (3) then their actual private signal was drawn depending on their harm type and the signal accuracy in that treatment; (4) finally, one member of each group was randomly selected and the payoffs of each member of that group depended on the *unconditional*

also examine these choices on an exploratory basis.

The decisions were presented in a fixed sequence. Participants first provided their decisions for when they received a no-harm signal (in Decision Screens 1–3) and then for when they received a high-harm signal (in Decision Screens 4–6). Within each set, the first decision involved the overwhelmed system (all others extract), the second decision involved the non-overwhelmed system (less than three others extracting), and the third decision was the unconditional choice. This ensured that the decision of interest always appeared as Decision 1 and the rest of the decisions followed in a logical order.

The decision screens, as well as the instructions and comprehension questions, varied depending on the treatment. These differences were clearly illustrated by the first decision screen in each treatment condition (*Accuracy*, *Cost*, and *Triage*) as compared to the *Baseline* (see Figure 1). *Accuracy* had different probabilities in point (ii): "Remember: There is a 90% chance that this message is true and a 10% chance that it is incorrect". *Cost* featured an additional message "Remember: You will incur an extraction fee of £0.50 if you are one of the group members who get to extract" directly below the decision question. Finally, *Triage* used a different framing of the choice: instead of choosing whether to extract or not, they were asked to report whether they were a high-harm type or a no-harm type. The decision screens of all treatments can be found in Appendix B.F, Screen 12.

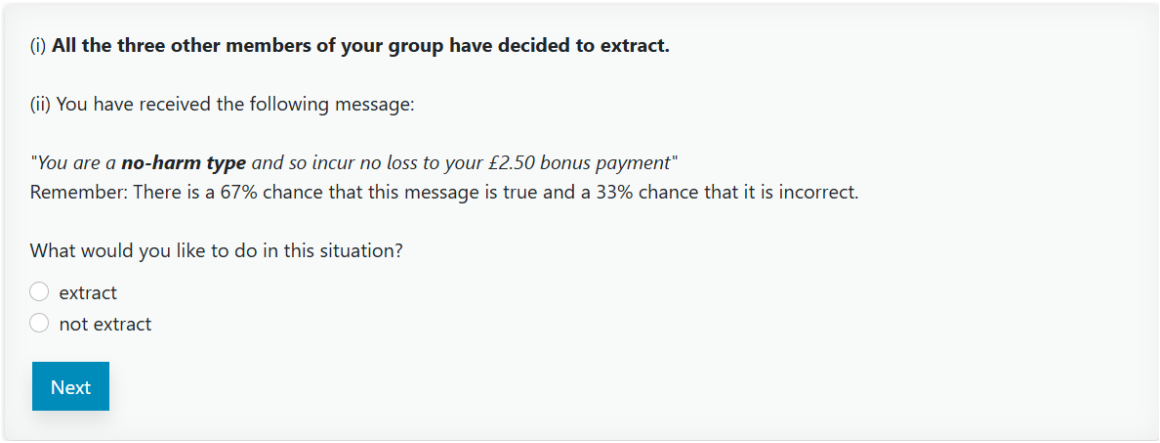


FIGURE 1: Decision screen for the baseline

We used Prolific's representative samples tool to gather a representative sample of the UK population in terms of gender, age, and ethnicity.[13] In total, 1696 UK subjects participated in two large sessions in the end of December 2024. Participants were randomly assigned to

choices of the *other* three group members and the corresponding *conditional* choice of that randomly selected group member, in each case depending on the actual harm type and private signal received by that group member.

[13]Prolific clarifies its representative sample tool as follows: *'When you use a representative sample, we take the intended sample size and stratify it across three demographics: age, sex and ethnicity. We use census data from the US Census Bureau or the UK Office of National Statistics to divide the sample into subgroups with the*

one of the four treatment conditions in each session. They made individual decisions online and were assigned to groups of four to calculate payoffs only after all the data had been collected. Individual observations can therefore be treated as statistically independent. The median duration was 11.6 minutes and the average payoff £3.43, which included a show-up fee of £1.50. Table 2 shows a summary of the overall demographics in the whole sample and in the treatments. Chi-squared tests reveal no significant differences in participant characteristics across treatments indicating that our randomization was successful.

TABLE 2: Summary Statistics by Treatment

Group	Overall	Baseline	Accuracy	Cost	Triage	p-value
Gender						0.843
Female	51.6%	50.0%	51.4%	51.9%	53.1%	
Male	48.4%	50.0%	48.6%	48.1%	46.9%	
Age Group						0.555
18-30	17.9%	20.1%	17.4%	17.4%	16.5%	
31-40	19.5%	19.4%	19.8%	17.1%	21.7%	
41-50	15.7%	15.5%	18.6%	15.5%	13.4%	
51-60	25.9%	25.7%	24.0%	26.4%	27.4%	
61+	21%	19.2%	20.2%	23.6%	21.0%	
Ethnicity						0.653
White	84.9%	82.9%	87.1%	84.3%	85.4%	
Asian	8.1%	9.3%	7.6%	7.9%	7.5%	
Black	3.4%	4.6%	2.4%	3.3%	3.1%	
Other	3.7%	3.2%	2.9%	4.5%	4.0%	
n	1696	432	420	420	424	

Note: Percentages may not add up to exactly 100% because of rounding. The p-values are based on Chi-squared tests for no differences in the composition of the groups.

III.C Results

In this section, we present the results of Experiment 1. Following our pre-registered hypotheses (see Appendix B.A for the pre-registration), we focus on behavior after a no-harm signal when all others extract (Section III.C.1), and report exploratory analyses for all other cases in Section III.C.2.

same proportions as the national population. This means, for example, that a representative sample contains the same proportion of 28-37 year old Asian women as the national population (or at least, as close as we can deliver)." Prolific (2025)

III.C.1 No-harm Signal

Figure 2 depicts the frequency of extraction by treatment after receiving a no-harm signal and under the condition that all others in their group have extracted. In each treatment, we observe that less than half of our sample chose to extract in this situation (47% in the Baseline). This suggests that at least some individuals care not only about maximizing their own payoffs but are also motivated by other factors, such as possibly the payoff of others, adhering to a social norm, or wanting to be seen as social. All these factors can be explained by a positive α in our model.[14]

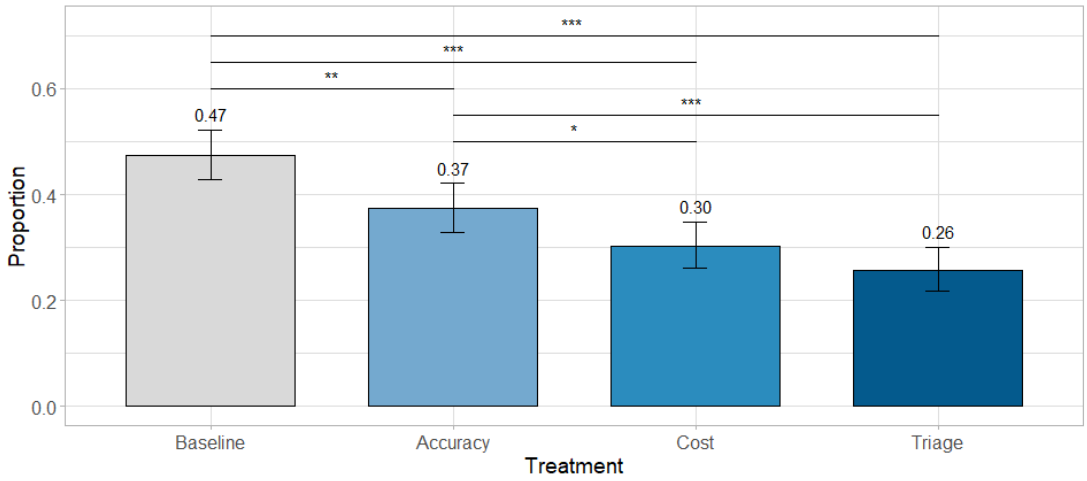


FIGURE 2: SHARE OF EXTRACTORS AFTER RECEIVING A NO-HARM SIGNAL UNDER THE CONDITION THAT ALL OTHERS EXTRACT. ERROR BARS REPRESENT WILSON CONFIDENCE INTERVALS. ALL SIGNIFICANCE LEVELS BASED ON TWO-SIDED PROPORTION TESTS: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We further observe that all three policy regimes lead to a significantly lower frequency of extraction relative to the *Baseline* (all at least $p < 0.01$, two-sided proportion tests). Additionally, the *Cost* and *Triage* treatment lead to a significantly lower frequency of extraction relative to the *Accuracy* treatment ($p < 0.05$ and $p < 0.001$ respectively, two-sided proportion tests). These results remain robust in a Probit regression with demographic controls, as well as in other robustness checks where we exclude participants with a less thorough understanding of the instructions (see Appendix B.B for details).

In relative terms compared to the Baseline extraction level, there is a 21% decrease in extraction frequency in the *Accuracy* treatment, a 36% decrease in the *Cost* treatment and a 45% decrease in the *Triage* treatment. Differences across treatments are even more pronounced

[14] Additional evidence for the existence of a positive α comes from the condition where not all others extract: in that case, extraction rates are significantly higher (see Section III.C.2).

for the unconditional extraction decisions after receiving a no-harm signal (see Figure 3 in Section III.C.2).

On exploring heterogeneity, we find that, overall, females extract less often than males ($p < 0.1$, Probit regression with robust SE, see Table 7 in Appendix B.B). We find no differences in the frequency of extraction depending on age or ethnicity, as well as no evidence for heterogeneous treatment effects based on any of these demographic variables (see Table 8 in Appendix B.B).

III.C.2 High-harm Signals, All Scenarios, and Efficiency

We complement our main analysis with exploratory analyses of extraction choices across all scenarios in which participants made a decision. Figure 3 provides an overview of the extraction shares in the four treatments depending on the signal received (no-harm in the top panel and high-harm in the lower panel) and depending on whether all others in one's group have extracted (left-most panel), not all others in one's group have extracted (middle panel) and finally, without any information about one's group members extraction choices (right-most panel).[15]

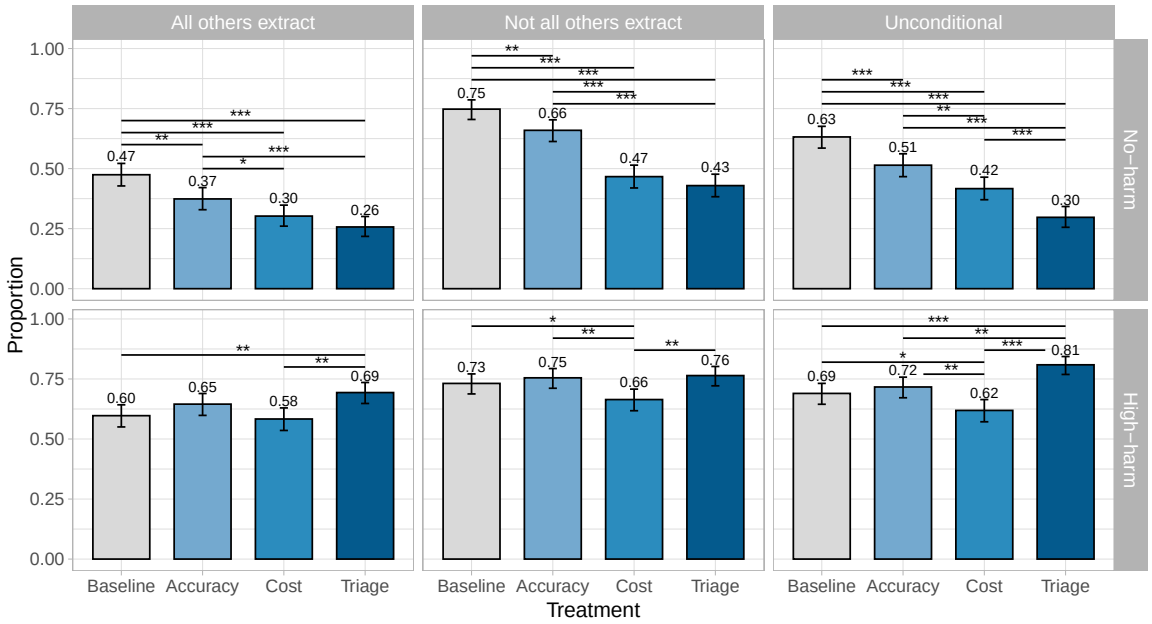


FIGURE 3: Overview of Share of Extractions per Signal and Condition. Error bars represent Wilson confidence intervals. Significance levels based on two-sided proportion tests: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[15]Though theory predicts that individuals should always extract after (i) receiving a high-harm signal or (ii) when not all others in one's group have extracted (i.e., assuming $\alpha < 1$), Figure 3 reveals that across all treatments, a substantial fraction of subjects refrain from doing so in these situations. Potential explanations for these deviations from theoretical predictions include an $\alpha > 1$, unaccounted costs such as norm violations, a general preference for non-extraction, or other out-of-the-lab norms and views.

The first takeaway is that after receiving a no-harm signal, all between-treatment differences are qualitatively similar and statistically significant both in the situation when not all others extract as well as when no information about group members' extraction choices is provided (i.e., the unconditional decision situation). This provides additional robustness to our main results. Second, we find that in the situation where not all other members in one's group have extracted, subjects extract significantly more ($p < 0.001$, Probit regression with robust and clustered SEs by individual, see Table 9 in Appendix B.D) than in the situation where all other members of one's group have extracted. This indicates that subjects do care about the externality of their extraction choices, i.e., it is evidence of $\alpha > 0$. Third, in accordance with theory, we find that extraction choices are significantly lower after receiving a no-harm signal compared to after receiving a high-harm signal ($p < 0.001$, Probit regression with robust and clustered SEs by individual, see Table 9 in Appendix B.D). This is evidence that subjects correctly interpret the nature of the signal and update beliefs about their type accordingly.

Turning to extraction choices after receiving a high-harm signal (lower panel of Figure 3), we find significantly higher extraction rates in the *Triage* treatment compared to the *Baseline* (at least $p < 0.01$, two-sided proportion tests) except under the condition that not all others extract, and significantly higher extraction rates in *Triage* compared to the *Cost* treatments across all decision situations (at least $p < 0.01$, two-sided proportion tests). Additionally, extraction rates are consistently lower in the *Cost* treatment compared to the *Baseline*, *Accuracy*, and *Triage* treatments (at least $p < 0.05$ in the unconditional decision situation as well as the situation in which not all members of one's group have extracted, two-sided proportion tests). We do not find differences between the *Accuracy* treatment and the *Baseline* treatment.

Finally, we evaluate the efficiency of the three institutional measures by computing the expected probability of excluding a high-harm player based on each individual's decision after receiving a no-harm signal and a high-harm signal (see Appendix B.E for details). We find that all treatments exhibit significantly higher efficiency compared to the Baseline (all $p < 0.001$, two-sided Wilcoxon rank sum tests). In the *Cost* and *Triage* treatments, this is a natural consequence of the reduced extraction by no-harm signal receivers in an overwhelmed system. In the *Accuracy* treatment there is an additional efficiency increase: due to less noisy no-harm signals, there are less no-harm signal receivers who end up being high-harm. This reduction in signal errors leads to the highest overall efficiency in the *Accuracy* treatment relative to all other treatments (at least $p < 0.01$, two-sided Wilcoxon rank sum tests). Efficiency in the *Triage* treatment is in turn higher than in the *Cost* treatment ($p < 0.01$, two-sided Wilcoxon rank sum tests) – driven by higher extraction rates in *Triage* compared to *Cost* after a high-harm signal.

In summary, we find that introducing a *Triage* system and/or an information intervention akin to the *Accuracy* treatment are preferable policy measures to instituting an upfront GP visit

cost, as they successfully reduce low-priority demand without the high-harm-signal deterrence observed under the *Cost* treatment.

IV Experiment 2: Vignette Experiment

Experiment 1 examines utilization in a supply-constrained system using an abstract game with neutral instructions, and reveals that all three proposed policy measures significantly reduce low-priority extraction from the common pool of resources. To complement these findings, in Experiment 2, we study the same dilemma in a less abstract setting, and examine the impact of the same three policy measures. In this new setting, rather than using induced monetary incentives, participants are asked to read a set of common health symptoms and report how likely they would be to visit their GP in each case.

IV.A Experimental Design

We employ five short vignettes to describe a set of common health symptoms and ask participants to indicate how likely they would be (on a scale of 0 to 100) to call their general practitioner's (GP's) office to schedule an appointment. These vignettes were developed in close consultation with medical professionals and then validated by 14 certified UK-based GPs (see Appendix C.D for more details). Before participants provided their response in each scenario, they read about how the hypothetical healthcare system works in their country. This description varies across four between-subject treatments. In the *Baseline*, participants are informed that they live in a country with universal health coverage and when needed, they can directly call their GP's office to schedule an appointment. GP visits are free and no out-of-pocket expenses are incurred for any medical procedures they may undergo. They are also made aware that the public healthcare system is financed through a fixed percentage of their monthly income collected by the state. To simulate the common pool resource dilemma from Experiment 1, we further explain that the primary healthcare system in their country is currently overwhelmed, with many patients not receiving timely care.

In the *Cost* treatment, they are informed that they pay a small upfront fee of £10 every time they visit their GP. The *Triage* treatment differs in the framing of the decision. Instead of calling to schedule an appointment with their GP, participants are informed: "To get an appointment at your local GP, you must first complete an online form describing your symptoms, and only patients who report severe symptoms will get an appointment. That said, you are free to exaggerate your symptoms to guarantee yourself an appointment." In this treatment, participants are therefore asked to indicate how likely they would be (on a scale from 0 to 100) to exaggerate their symptoms to obtain an appointment. To examine the interaction between

the *Cost* and *Triage* treatments, the experiment also includes a *CostTriage* treatment, wherein they both pay the £10 fee for every GP visit and need to exaggerate their symptoms in order to secure an appointment.

In all treatments, participants are presented with five scenarios. Scenarios A, B, C, and D are low-*harm* scenarios, mirroring the no-harm signal in Experiment 1. The described symptoms in these scenarios are relatively mild and unlikely to warrant a GP visit, but as with any set of health symptoms, there is a low probability they could indicate a serious or more long-lasting adverse health condition that does warrant medical attention (the stochastic harm, A, from Experiment 1). Scenario E describes symptoms highly likely to warrant a GP visit, and is thus equivalent to receiving a high-harm signal in Experiment 1. The validation of the scenarios by the 14 certified UK GPs was aimed at ensuring that the low-harm scenarios would not require a GP visit, while the high-harm scenario would warrant one. To manipulate signal precision, each scenario had two versions. A shorter version, which we call the low-accuracy version, and a version with additional information about the symptoms, typically one or two sentences, which we call the high-accuracy version. Figure 4 displays the decision screen for Scenario A in the *Baseline* treatment, with the sentence starting with "You remember" added in the high-accuracy scenario.

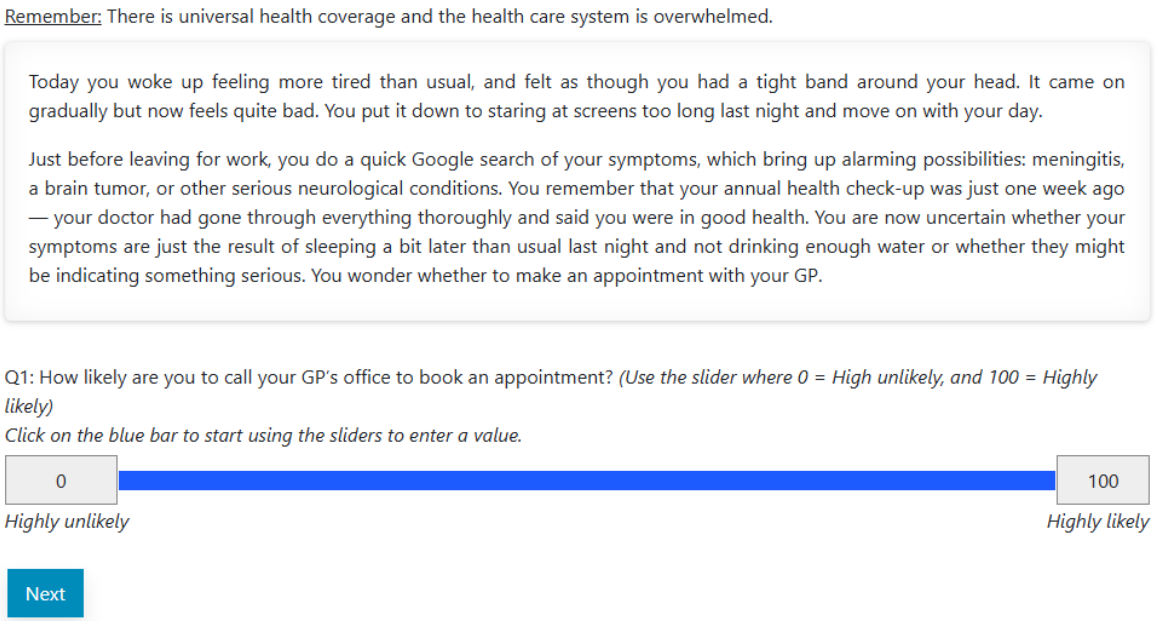


FIGURE 4: DECISION SCREEN FOR THE *BASELINE* TREATMENT

Scenarios B and C follow an analogous structure: each describes mild symptoms (e.g., cramp-like lower abdominal pain or a mild cough) combined with the contemplation of a highly unlikely but serious diagnosis, such as appendicitis or lung cancer. The high-accuracy version of Scenario B (crampy pain) adds: "You also recall that you barely drank any water yesterday

and that you have actually experienced similar symptoms once before and it passed on its own." The high-accuracy version of Scenario C adds: "You force yourself to be calm and think through things. You remember at your last health check a few weeks ago, the GP said you were doing fine and they were quite thorough with the tests. Further googling suggests that a mild cough is rarely indicative of a serious health issue." In Scenario D, the symptoms consist of a sore throat likely caused by a viral infection. In the high-accuracy version, the additional information states: "You realize that you have most likely caught the viral infection your niece had and in that case, you know that nothing much can be done except to wait it out." Finally, Scenario E presents symptoms consistent with a potential heart issue (nausea, a dull ache in the upper back, and shortness of breath). In the high-accuracy version, a friend who is a medical professional recommends an immediate visit to one's local physician. The full text of all scenarios, with high-accuracy variations clearly indicated, are provided in Appendix C.G.

We examine the interaction between the four between-subject treatments (*Baseline*, *Cost*, *Triage*, and *CostTriage*) and *Accuracy* by varying whether participants view the low or high-accuracy version of a given scenario in all treatments. Each participant is randomly assigned to one of two scenario combinations. In combination T1, Scenarios A and C are presented in their high-accuracy versions; in combination T2, Scenarios B, D, and E, appear in the high-accuracy version. The order of the five scenarios is randomized, but identical across matched pairs of T1 and T2 observations to ensure maximum comparability. The accuracy of the signal is varied within-subject to reduce potential boredom or need to be consistent that may have come into play had all scenarios appeared overly similar. At the same time, the scenario combinations are varied between-subject to ensure that each scenario is observed in both its low- and high-accuracy version.

IV.B Predictions

Our objective in Experiment 2 is to align the incentive structure across treatments as closely as possible with that of Experiment 1, while simultaneously ensuring that the decision context remains similar to an actual health situation an individual might experience. Table 3 compares the main features of the two experiments.

Some features explicitly modeled in Experiment 1, such as harm A, accuracy k, and the extraction externality, are conveyed more indirectly in Experiment 2 through information about the symptoms experienced and the state of the hypothetical healthcare system respectively.

As it concerns the between-subject treatments, in Experiment 1, the *Cost* treatment features an extraction fee amounting to £0.50. In Experiment 2, there is likely to be a perceived non-monetary hassle cost of visiting one's GP (not captured in Experiment 1), which is then supplemented by a *hypothetical* co-payment of £10.00 in the *Cost* treatments. In the *Triage*

treatments, meanwhile, while a psychic cost (c_i^{psych}) is imposed in both experiments, it differs in its framing. In Experiment 1, this psychic cost is incurred because extraction requires reporting one is a high-harm type after receiving a no-harm signal. This also means subjects who receive a high-harm signal cannot incur this cost. In Experiment 2, this cost represents participants' potential disutility from having to exaggerate their symptoms to obtain a GP appointment. While this cost could be incurred both after receiving a low-harm signal (Scenarios A–D) as well as a high-harm signal (Scenario E), exaggeration itself would be less necessary in Scenario E where the described symptoms are already quite serious.

TABLE 3: Comparison of experiments

	Laboratory experiment (experiment 1)	Vignette experiment (experiment 2)
Harm A	Monetary cost of £2.50	Subjective assessment of the severity of the health scenarios
Accuracy k	– 2/3 in *Baseline*, *Cost*, and *Triage* – 9/10 in *Accuracy*	– Subjective assessment of the health scenarios in *Low Accuracy* – Increased through additional information in *High Accuracy*
Extraction cost c_i	– £0.00 in *Baseline*, *Triage*, and *Accuracy* – £0.50 in *Cost*	– Perceived hassle cost of visiting a GP in *Baseline* and *Triage* – Additional hypothetical co-payment of £10.00 in *Cost* and *CostTriage*
Psychic cost c_i^{psych}	– No psychic lying cost of extraction in *Baseline*, *Accuracy*, and *Cost* – Extraction requires reporting being a high-harm type after receiving a no-harm signal in *Triage*	– No psychic lying cost of a GP visit in *Baseline* and *Cost* – Obtaining a GP visit requires exaggerating one's symptoms in *Triage* and *CostTriage*
Extraction externality	Implemented through limited capacity, where only $n-1$ players can extract	Conveyed through information that the primary healthcare system is overwhelmed, resulting in many patients not receiving timely care

We rely on the comparative statics of our theoretical model to derive qualitative preregistered predictions for Scenarios A–D in Experiment 2 (i.e., where a low-harm signal is received).[16] These predictions are summarized in Table 4. As in Experiment 1, we predict that each individual treatment (*Cost*, *Triage* and *Accuracy*) reduces the likelihood of extraction (see Predictions 1, 2, and 6). In addition, the design of Experiment 2 allows us to test interaction

[16]The model can be readily extended to include a low-harm instead of a no-harm signal as long as the associated loss is lower than the high-harm loss, i.e. $B < A$. The decision tradeoff (Equation 2) becomes the following:

$$\underbrace{\frac{n-1}{n}(p_h(h,\sigma,k)A + (1-p_h(h,\sigma,k))B)}_{\text{Benefits of decision to extract}} > \underbrace{\alpha_i \frac{n-1}{n}(hA+(1-h)B)}_{\text{Altruistic cost}} + \underbrace{\frac{n-1}{n}c_i}_{\text{Extraction cost}} + \underbrace{c_i^{psych}\mathbb{1}_{Triage \wedge \sigma_{no}}}_{\text{Lying cost}}$$

All comparative statics about the policy interventions remain qualitatively the same, even if subjects do not consider the externality of their decision (i.e., if $\alpha = 0$).

effects. The interaction between *Cost* and *Triage* is identified through the between-subject treatment *CostTriage*. The interactions between *Accuracy* and *Cost*, *Triage*, and *CostTriage* are identified using both within- and between-subject comparisons of the low- and high-accuracy versions of the four low-harm scenarios. In Predictions 3–5 and 7–9, we hypothesize that all interaction effects will further reduce the likelihood of extraction relative to any individual treatment alone.

TABLE 4: Overview of pre-registered predictions in Experiment 2 for Scenarios A–D

Prediction	Comparison of the extraction likelihood in treatments
Prediction 1	*Baseline* > *Cost*
Prediction 2	*Baseline* > *Triage*
Prediction 3	*Baseline* > *CostTriage*
Prediction 4	*Cost* > *CostTriage*
Prediction 5	*Triage* > *CostTriage*
Prediction 6	*Low Accuracy* > *High Accuracy* in *Baseline*
Prediction 7	*Low Accuracy* > *High Accuracy* in *Cost*
Prediction 8	*Low Accuracy* > *High Accuracy* in *Triage*
Prediction 9	*Low Accuracy* > *High Accuracy* in *CostTriage*

IV.C Experimental Procedures

Experiment 2 was programmed in oTree (Chen et al., 2016) and conducted online via Prolific.[17] Participants first read the experimental instructions (see Appendix C.F), which included a description of the hypothetical healthcare setting and three comprehension questions. As in Experiment 1, participants who answered a comprehension question incorrectly received a brief explanation and were required to try again until they provided the correct response. Participants then proceeded to read the five health scenarios. For each scenario, they indicated the likelihood (on a scale from 0 to 100) that they would call the GP's office (*Baseline*, *Cost*) or, in the *Triage* and *CostTriage* treatments, the likelihood that they would exaggerate their symptoms in an online form to obtain a GP appointment. Finally, participants completed a short post-experimental questionnaire collecting socio-economic characteristics. The questionnaire also included an open-text question about the reasons behind their choices in the five scenarios.

As in Experiment 1, we used Prolific's representative samples tool to gather a representative sample of the UK population in terms of gender, age, and ethnicity. Those, who already participated in Experiment 1 were excluded in the sample selection. In total, we collected data from 1200 UK subjects. The median duration was approximately 5 minutes and everyone received

[17]The experiment was pre-registered at https://aspredicted.org/dh7mb6.pdf and received approval from the Research Ethics Committee at the University of Reading.

the same completion payment of £1.00. Table 2 provides a summary of the demographics in the sample by treatment. Chi-squared tests reveal no significant differences in participant characteristics across treatments indicating successful randomization.

TABLE 5: SUMMARY STATISTICS BY TREATMENT

Group	Overall	Baseline	Cost	Triage	CostTriage	p-value
Gender						$p = 0.127$
Female	51.7%	46.9%	53.2%	56.2%	50.5%	
Male	48.3%	53.1%	46.8%	43.8%	49.5%	
Age Group						$p = 0.654$
18-30	18.1%	14.4%	21.1%	17.5%	19.4%	
31-40	19.4%	19.3%	19.4%	18.2%	20.7%	
41-50	15.8%	16.1%	14.0%	18.2%	15.1%	
51-60	24.4%	25.6%	26.1%	22.6%	23.4%	
60+	22.2%	24.6%	19.4%	23.6%	21.4%	
Ethnicity						$p = 0.224$
White	85.3%	86.9%	83.9%	83.8%	86.6%	
Asian	7.8%	8.5%	6.7%	9.1%	6.7%	
Black	3.3%	2.3%	4.0%	4.7%	2.3%	
Other	3.6%	2.3%	5.4%	2.4%	4.3%	
n	1200	305	299	297	299	

Note: Percentages may not add up to exactly 100% because of rounding. The p-values are based on Chi-squared tests for no differences in the composition of the groups.

IV.D Results

Next, we present the results of Experiment 2, closely following our pre-analysis plan (see Appendix C.A for the pre-registration). Again, our main focus is on decisions after receiving a low-harm signal, i.e., in Scenarios A–D (Section IV.D.1). This is complemented with exploratory analyses of choices in each of the individual low-harm scenarios as well as in the high-harm signal scenario (Scenario E) in Section IV.D.2.

IV.D.1 Low-harm Signals

In this Section, we analyze treatment effects in the four scenarios where a GP visit is not warranted, i.e., our low-harm scenarios, by running the following two regressions:

Regression I examines differences across our between-subject treatments i.e., *Cost*, *Triage*, and *CostTriage*:

$$y_{is} = \alpha + \beta' Treat_i + \delta' ScenarioAccuracy_s + \epsilon_{is}$$

where y_{is} denotes the reported likelihood of scheduling a GP appointment by individual i in scenario s. $Treat_i$ is a vector of between-subject treatment dummies (*Cost*, *Triage*, and *CostTriage*). $ScenarioAccuracy_s$ is a vector of seven dummies, which encompass fixed effects for each scenario and its accuracy version. The model is estimated via OLS with robust standard errors clustered at the individual (i) level.

Regression II assesses whether receiving a more accurate signal about one's health status affects the likelihood of scheduling an appointment in each treatment:

$$y_{is} = \alpha + \beta' Treat_i + \gamma'(Treat_i^{+Baseline} \times High - accuracy_{is}) + \delta' Scenario_s + \epsilon_{is}$$

where y_{is} and $Treat_i$ are as defined above; $Treat_i^{+Baseline}$ is a vector of dummies for all treatments, including for the Baseline, $High - accuracy_{is}$ is a vector of dummies indicating whether scenario s is the high- or low-accuracy version; and $Scenario_s$ are scenario fixed effects pooled across both accuracy conditions. The model is estimated via OLS, with robust standard errors clustered at the individual (i) level.

Table 6 shows the estimation results of Regression I and Regression II. In both specifications, we do not find evidence for any significant effect of the *Cost*, *Triage*, or *CostTriage* treatments on the reported likelihood of scheduling a GP appointment ($p = 0.75$, $p = 0.62$, $p = 0.83$, respectively, OLS with robust and clustered SE). All estimated coefficients are close to zero in magnitude (Cohen's d $\approx$ -0.02 to 0.04). Thus, we do not replicate the effects of the *Cost* and *Triage* treatments from Experiment 1. We also do not find any evidence for differences between these treatments ($p = 0.61$, joint F-based Wald test).

However, we find that viewing the high-accuracy version of a given scenario leads to a significant reduction in the reported likelihood of scheduling a GP appointment in all our between-subject treatments ($p < 0.001$ for *Cost*, *Triage*, *CostTriage*, OLS with robust and clustered SEs). This is consistent with the effect of the *Accuracy* treatment from Experiment 1. Furthermore, the results remain robust when restricting the analysis to the first displayed scenario for each individual, where the variation in accuracy is purely between subjects as the first displayed scenario was randomly determined (see Table 10 in Appendix C.B).[18]

Although the coefficients of all interactions of the treatments with the high-accuracy dummies differ slightly, we find no evidence that these interaction effects differ systematically across treatments ($p = 0.94$, joint F-based Wald test).

[18]In this specification, the interactions between high accuracy and the *Cost*, *Triage*, and *CostTriage* treatments are larger in magnitude and statistically significant ($p < 0.001$, $p < 0.05$) or marginally significant ($p < 0.1$). We also observe a significant ($p < 0.05$) *increase* in the reported likelihood of scheduling a GP appointment in the *Triage* treatment relative to the *Baseline* – different to the absence of an effect in the main specification, which focuses on all scenarios. Given that the analysis is based on a subset of decisions and arises from an exploratory analysis, this increase should be interpreted with caution.

TABLE 6: Regression Results (Experiment 2)

	Reported likelihood of scheduling a GP appointment (%)	
	Regression I	Regression II
	(1)	(2)
Cost	−0.459	−0.618
	(1.463)	(1.671)
Triage	0.777	0.488
	(1.556)	(1.699)
CostTriage	0.831	0.435
	(1.450)	(1.615)
Baseline × High accuracy		−3.855***
		(0.939)
Cost × High accuracy		−3.538***
		(0.994)
Triage × High accuracy		−3.301***
		(0.826)
CostTriage × High accuracy		−3.099***
		(0.876)
Constant	20.242***	18.255***
	(1.355)	(1.236)
Scenario Fixed	No	Yes
Scenario Accuracy Fixed	Yes	No
#Clusters (Individuals)	1,200	1,200
Observations	4,800	4,800
R^2	0.012	0.009
Adjusted R^2	0.010	0.007

*Note: Robust SEs clustered by participant. *p<0.05; **p<0.01; ***p<0.001*
In Regression I, the constant captures the mean outcome in the the low-accuracy version of Scenario C in the Baseline condition while in Regression II, it is the mean outcome in Scenario C in the Baseline.

The interaction between treatment assignment and high-accuracy reduces scheduling likelihood by approximately 3–4 %-points across all treatments. Relative to a baseline mean of about 18%, this corresponds to a reduction of roughly 17–21%. Although the standardized effect size is modest when measured against the overall dispersion of the outcome (Cohen's d ≈ -0.15), the magnitude is still substantively relevant in an overcrowded system, where even a moderate reduction in unnecessary GP visits would allow for more efficient allocation of medical resources.

On examining heterogeneity in our sample, we find that, overall, females, ethnically white participants, and participants over 50 years old are significantly less likely to schedule a GP appointment ($p < 0.001$, $p < 0.001$, and at least $p < 0.01$, respectively; OLS with robust and clustered SE, see Table 11 in Appendix C.C). However, we find no evidence of heterogeneous treatment effects in Regression I or II (see Table 12 in Appendix C.C). The main results from above are robust to the inclusion of these demographics.

IV.D.2 High-harm Signal and Individual Health Scenarios

Figure 5 summarizes the reported likelihood of scheduling a GP appointment across all five health scenarios, separately by treatment and by accuracy version. The full wording of all scenarios in both accuracy versions is provided in Appendix C.F.

The mean reported likelihood of scheduling a GP visit in the four low-harm scenarios ranges from 13-21%. The upper row of Figure 5 shows the reported scheduling likelihood averaged across low- and high-accuracy versions of a given scenario for each of the *Cost*, *Triage*, and *CostTriage* treatments. In the four low-harm scenarios, reported likelihoods are very similar across treatments. Only Scenario A exhibits slightly higher average reports and slightly larger dispersion across treatments.

The lower row of Figure 5 shows the reported likelihood of scheduling a GP visit conditional on the accuracy condition.[19] We find a significant reduction (at least $p < 0.05$, two-sided Wilcoxon Rank-Sum Test) in this reported likelihood under the high-accuracy versions of the scenarios for all except Scenario B.[20]

[19]Note that each participant saw each scenario only in one of the two accuracy versions, making the scenario accuracy comparisons between-subject.

[20]It is possible we do not observe an effect of higher accuracy in Scenario B because, compared to the other three low-harm scenarios, the high-accuracy version of Scenario B does not provide as much relevant information relative to its low-accuracy counterpart. For instance, the high-accuracy version of Scenarios A and C both refer to a clean bill of health from a recent annual health check (very relevant if debating whether to visit the GP today), while Scenario D gives a highly plausible alternate explanation for one's symptoms, which if true, would definitely *not* call for a GP visit. By contrast, in Scenario B, the high accuracy version only refers to being potentially dehydrated and the possibility that the symptoms would pass on their own, which perhaps is insufficient information to reliably shift the perceived benefits of visiting one's GP today.

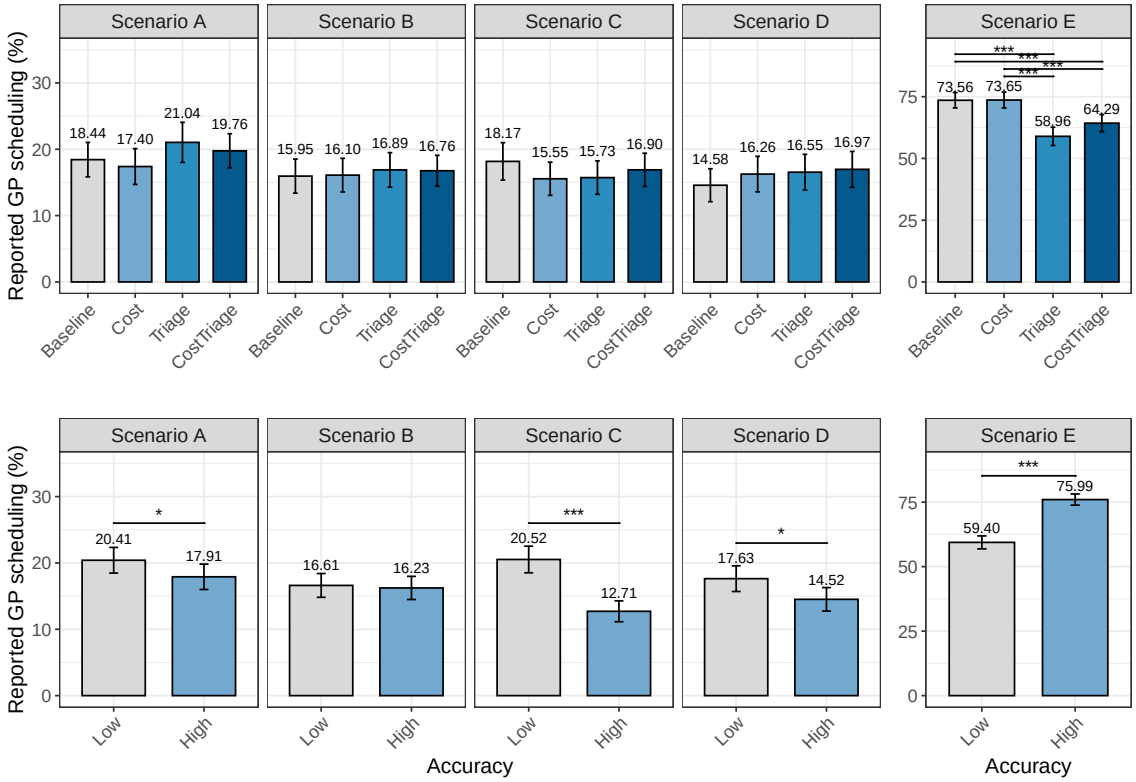


FIGURE 5: Overview reported likelihood of scheduling a GP appointment per scenario and treatments. *** $p < 0.001$, * $p < 0.05$, based on two-sided Wilcoxon Rank-Sum Tests.

We next turn to the high-harm Scenario E,[21] shown on the right-hand side of Figure 5. The first observation is that mean likelihoods to schedule an appointment are around 60-75% and thus substantially higher than in the low-harm scenarios.

Comparing treatments, we find no evidence that *Cost* affects the reported likelihood of scheduling an appointment ($p = 0.84$, two-sided Wilcoxon Rank-Sum Test). We observe a significant reduction in reported likelihoods in the two *Triage* treatments (all $p < 0.001$, two-sided Wilcoxon Rank-Sum Tests) but, as pre-registered, do not interpret these effects (see Appendix C.A). This is because exaggerating symptoms may be perceived as unnecessary when symptoms are already severe, thus mechanically reducing responses to the question on exaggeration.[22] Importantly, we find that viewing the high accuracy version of the high-harm scenario leads to a significant increase in the reported likelihood to visit one's GP pooled over all treatments ($p < 0.001$, two-sided Wilcoxon Rank-Sum Tests) as well as in all three treatments

[21] In the validation study, all GPs fully agreed that a doctor should be consulted in Scenario E (see Appendix C.D).

[22] This interpretation is consistent with at least one of the participants' comments: *"No point exaggerating my symptoms regarding No. 5 as I think my symptoms would be enough to alert the doctor but I would be detailed about them and express my concerns. (...)"*.

separately (at least $p < 0.01$, two-sided Wilcoxon Rank-Sum Tests). This increase amounts to approximately 16%-points, underscoring the importance of ensuring that all individuals, and in particular those experiencing serious symptoms, have access to more accurate information about the symptoms they are experiencing.

V Discussion

We conducted two experiments to study the effect of different policy regimes on managing excess demand for limited health resources. The two studies examine the same choice under different conditions: Experiment 1 has a well-defined set of parameters and elicits the main decision using monetary incentives, while Experiment 2 simulates the same type of trade-offs in a more general health-framed framework that leaves room for individual interpretation about the state of the healthcare system and the consequences of visiting one's GP. This two-pronged approach enables us to test our hypotheses in two different contexts and improves the generalizability of the finding that better signal precision can effectively reduce low-priority demand and improve overall efficiency. At the same time, the scope and strength of each of the treatments we employed may have varied across the two settings. In this section, we discuss features of both settings that may help explain the observed differences and the applicability of our findings to real healthcare systems.

A central difference between the two experiments lies in the nature and magnitude of the costs and benefits associated with seeking care. In Experiment 2, both are hypothetical, but the potential consequences – such as failing to detect a potentially life-threatening condition such as cancer – are severe. While such outcomes cannot be credibly incentivized, their salience nevertheless shapes behavior. In particular, the prospect of a severe health loss may direct participants' attention primarily toward their symptoms, making them less responsive to monetary costs or the necessity to overstate one's symptoms.[23] Consistent with this interpretation, in a post-experimental questionnaire, most participants justify their decisions in Experiment 2 primarily by referring to the described symptoms across the different scenarios.[24]

[23] Additionally, participants may not fully integrate costs across domains. In health-framed decisions, the possibility of severe health losses may dominate relatively minor monetary or psychological costs, thereby attenuating the effectiveness of the *Cost* and *Triage* treatments in Experiment 2. This interpretation is consistent with models in which health and consumption are not additively separable, so that the marginal utility of money may depend on health status (Finkelstein et al., 2013).

[24] Across all treatments, text-analysis reveals that approximately 70% of participants justify their decision to schedule (or not schedule) a GP visit by referring to symptom-related considerations (e.g., symptom, severity, urgency, ill), with the word *symptom* being the most frequently used term across all treatments. Only about 10% additionally cite system-related concerns, e.g., the overwhelmed system, not wanting to burden the GP or the healthcare system. In the *Cost* treatments, roughly 10% mention the monetary cost, often noting, however, that it did not affect their decision. Similarly, in the *Triage* treatments, only around 10% refer to the need to exaggerate symptoms as part (or not part) of their decision-making. See Appendix C.E for details.

In Experiment 1, incentives are real but moderate in magnitude. Although the potential monetary loss associated with being a high-harm type is meaningful in the context of a short online experiment, it is arguably modest relative to the possibility of a severe health consequence – a small but distinct risk present in each of the no-harm scenarios of Experiment 2. Moreover, because all aspects of extraction are explicitly defined in monetary terms in Experiment 1, participants can more readily weigh multiple considerations, such as extraction costs and potential negative externalities, in a balanced manner. This likely increased the effectiveness of some of the treatment interventions relative to Experiment 2.

Coming to the specific treatments, the relative magnitude and nature of the *Cost* intervention differed across the two settings. In Experiment 1, participants faced a real extraction cost of £0.50 that represented a meaningful component of the overall payoff structure and likely entered the decision process directly. In Experiment 2, participants instead faced a hypothetical £10 co-payment embedded in a broader health context involving potentially severe consequences. Although nominally larger, its effect may have been attenuated by its hypothetical nature and because participants evaluated it relative to perceived health risks rather than as a purely monetary trade-off. This suggests that the effectiveness of financial interventions may depend not only on their absolute size, but also on how salient they are relative to other aspects of the decision environment.

A related difference arises in the *Triage* treatment, where the intervention itself may be perceived differently across the two settings. In Experiment 1, accessing the common pool in the *Triage* treatment requires dishonest reporting, potentially triggering lying aversion. In Experiment 2, it involves stating a willingness to exaggerate one's symptoms in an online triage form. While similar in spirit, such forms may be perceived as a routine feature of accessing care. As a result, the intervention may be interpreted as a procedural hurdle rather than a moral cost, weakening its impact on behavior. Importantly, this need not merely reflect a difference in experimental implementation, but may capture a realistic feature of healthcare systems, where interventions intended to discourage low-priority demand can become normalized and therefore lose part of their intended effect.

Finally, participants may evaluate the private benefits of seeking care differently across the two settings. In Experiment 2, consistent with behavioral hazard in health decisions under full insurance (Baicker et al., 2015), individuals might overestimate the benefit of seeing their GP. Indeed, the four low-harm scenarios in this experiment describe concrete and familiar symptoms which may focus attention on their personal experience of these symptoms rather than on any social cost of visiting their GP. Providing more accurate information about the severity of their condition therefore reduces intended GP visits by helping participants recalibrate the expected benefit of seeking care. In Experiment 1, signal accuracy is presented as an abstract

numerical parameter, which may not be as salient when considering whether or not to extract. Nonetheless, we continue to observe a reduction in extraction behavior, suggesting that improving information may remain effective even when presented in an abstract way.

Taken together, the results from our two experiments indicate that the effectiveness of policies designed to curb low-priority demand depends critically on how the decision is framed. When the externality associated with extraction is salient and the consequences of non-extraction are clearly defined in monetary terms – as in the abstract setting of Experiment 1 – introducing small monetary or psychological misreporting costs can discourage low-priority utilization. However, when individuals evaluate the decision primarily through the lens of their own health and the symptoms they are experiencing – as in Experiment 2 – such interventions appear to have little influence. Instead, in that context, we observe that participants focus on the perceived personal benefit of seeking care, and only interventions that meaningfully shift these perceived benefits reduce the propensity to schedule a GP visit. Across both settings, however, interventions that improve the accuracy of information about one's potential harm remain effective, as they directly shape individuals' perceptions of whether care is warranted. This suggests that information-based interventions may capture a comparatively robust mechanism whose effectiveness is less dependent on the particular decision environment and may therefore be more likely to translate to real healthcare settings.

VI Conclusion

We conduct two large online experiments with representative samples of the UK population to investigate the impact of three policy interventions on managing the excess demand for primary healthcare services. Combining a laboratory-style design (Experiment 1) with a complementary health-framed survey experiment (Experiment 2), we study how individuals with uncertain health needs and noisy signals respond to different institutional features. Across both experiments, we find consistent evidence that providing more accurate information about one's potential harm reduces the unnecessary extraction of resources. By contrast, while both an extraction fee and the requirement to overstate symptoms reduce extraction in the abstract setting of Experiment 1, these effects do not replicate in the health-framed environment of Experiment 2.

These findings suggest that policies aimed at improving individuals' information about their health condition before they decide to seek out care from a health professional are likely to be the most robust tool for reducing overuse. Such interventions enable a form of self-triage, allowing patients to better assess their needs before entering the system, thus alleviating the demand-side pressure on the health-care system. This may be particularly valuable in health-care systems where increasing supply is not possible in the short run.

The divergence between some of our laboratory and vignette results also carries a methodological implication. When studying behavior in experimental settings designed to capture the incentives of key actors in the health system, our findings suggest researchers should consider complementing incentivized laboratory-style designs with health-framed survey experiments. This is particularly relevant for patient behavior, where field experiments are often infeasible or ethically constrained, and where realistic incentivization in the laboratory is difficult. Combining both approaches can therefore provide a more accurate approximation of real-world behavior than relying on either method in isolation.

The main limitation of the current study is that it does not touch on the supply-side of the problem in primary care. Interventions aimed at improving efficiency, e.g. through informative triage, improving continuity of care or considering the incentives of medical personnel themselves, are outside the scope of the current study, but could be very effective at addressing wait times, increasing access and improving overall health outcomes. Hence, information provision should be viewed as a complement rather than a substitute for other policy tools. In particular, combining improved patient-side information with effective supply-side triage, investments in capacity, and preventive care may yield the largest welfare gains.

Another limitation of our study is that it focuses primarily on settings characterized by potential overuse. Our measure of efficiency therefore largely reflects reductions in low-priority care. While we find that improved information increases extraction among individuals in the case of our high-harm scenario – suggesting that such interventions can also promote appropriate care-seeking when needs are acute – we do not specifically cover scenarios that describe symptoms, where patients typically under-utilize primary care resources, such as in cases of ADHD or depression. An important direction for future research could be to examine the effectiveness of different interventions in the case of health conditions where under- (rather than over-) utilization is more prevalent.

Overall, our findings highlight that improving individuals' information about their health needs is a simple yet effective tool for reducing low-priority demand while preserving access for those who need care. By targeting informational frictions on the demand side, such interventions may offer a scalable approach to improving the efficiency of primary care systems. In contrast, interventions that rely on financial costs, such as co-payments, or on behavioral frictions, such as symptom exaggeration requirements imposed by supply-side triage, appear more sensitive to context and may not translate as reliably to real-world health-related behavior.

Acknowledgments

We thank Aleksandra Khokhlova for providing valuable input on the text analysis of participant responses in Experiment 2. We thank Dr. Sandhya Abraham for her valuable feedback

and comments on the health scenarios used in Experiment 2. We thank Steven Bosworth, Kamila Cygan-Rehm, Pierre Dubois, Johanna Kokot, Gergely Hajdu, Glenn Harrison, Aleksandra Khokhlova, Wieland Müller, Simon Reif, and Rostislav Staněk for helpful comments and discussions. We also appreciate comments from participants at the 15th Annual Conference of the French Association of Experimental Economics (ASFEE) in Nancy, the 2025 ESA European Meeting in Brno, the 17th Nordic Conference in Behavioral and Experimental Economics in Stavanger, the 10th Workshop in Behavioral and Experimental Health Economics in Rotterdam, internal research seminars at Masaryk University and the University of Reading, as well as from participants from the Economics & Sustainability Seminar at the University of Regensburg, the IMS/CoBe Research Meeting at WU Vienna and the BEEL Experimental Economics Workshop in Birmingham. This output was supported by the NPO "Systemic Risk Institute" number LX22NPO5101, funded by European Union - Next Generation EU (Ministry of Education, Youth and Sports, NPO: EXCELES).

References

Abeler, J., Becker, A., and Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, 113:96–104.

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753.

Aron-Dine, A., Einav, L., and Finkelstein, A. (2013). The rand health insurance experiment, three decades later. *Journal of Economic Perspectives*, 27(1):197–222.

Baicker, K., Mullainathan, S., and Schwartzstein, J. (2015). Behavioral hazard in health insurance. *The Quarterly Journal of Economics*, 130(4):1623–1667.

Baji, P., Boncz, I., Jenei, G., and Gulácsi, L. (2010). The short story of co-payments for health care services in Hungary–lessons for neighbouring countries. *Zdrowie Publiczne i Zarządzanie*, 8(1).

Batavanis, M., Arora, A., Aryee, S., and Nathwani, N. (2025). Exploring barriers to expanding medical training numbers in england: a national survey of medical education directors. *Clinical Medicine*, page 100523.

Bauer, T., Augurzky, B., and Schaffner, S. (2006). Copayments in the german health system-does it work? Technical report, CEPR Discussion Papers.

Blunt, I., Edwards, N., and Merry, L. (2015). What's behind the a&e 'crisis'. *London: Nuffield Trust*.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Kokot, J., and Wiesen, D. (2024). A new look at physicians' responses to financial incentives: Quality of care, practice characteristics, and motivations. *Journal of Health Economics*, 94:102862.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., and Wiesen, D. (2017). The effects of introducing mixed payment systems for physicians: Experimental evidence. *Health Economics*, 26(2):243–262.

Çakıcı, Ö. E. and Mills, A. F. (2021). On the role of teletriage in healthcare demand management. *Manufacturing & Service Operations Management*, 23(6):1483–1504.

Care Quality Commission (2024). The state of health care and adult social care in england 2023/24. Technical report, Care Quality Commission, London, UK. Accessed January 2026; shows GP registration increase of 10% between 2016–2024 and a 15% reduction in GPs per 100,000 patients in England.

Chandra, A., Flack, E., and Obermeyer, Z. (2024). The health costs of cost sharing. *The Quarterly Journal of Economics*, 139(4):2037–2082.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chiappori, P.-A., Durand, F., and Geoffard, P.-Y. (1998). Moral hazard and the demand for physician services: first lessons from a french natural experiment. *European Economic Review*, 42(3-5):499–511.

Cockx, B. and Brasseur, C. (2003). The demand for physician services: Evidence from a natural experiment. *Journal of Health economics*, 22(6):881–913.

Cowling, T. E., Harris, M. J., Watt, H. C., Gibbons, D. C., and Majeed, A. (2014). Access to general practice and visits to accident and emergency departments in England: cross-sectional analysis of a national patient survey. *The British Journal of General Practice*, 64(624):e434.

Dardanoni, V. and Wagstaff, A. (1990). Uncertainty and the demand for medical care. *Journal of Health Economics*, 9(1):23–38.

Dolton, P. and Pathania, V. (2016). Can increased primary care access reduce demand for emergency care? evidence from england's 7-day gp opening. *Journal of Health Economics*, 49:193–208.

Estrada Rodriguez, A., Ibrahim, R., and Zhan, D. (2025). On customer (dis-) honesty in unobservable queues: The role of lying aversion. *Management Science*, 71(1):844–860.

Finkelstein, A., Hendren, N., and Luttmer, E. F. (2019). The value of medicaid: Interpreting results from the oregon health insurance experiment. *Journal of Political Economy*, 127(6):2836–2874.

Finkelstein, A., Luttmer, E. F., and Notowidigdo, M. J. (2013). What good is wealth without health? the effect of health on the marginal utility of consumption. *Journal of the European Economic Association*, 11(suppl_1):221–258.

Fisher, B. (2026). Why does england have unemployed gps when patients can't get gp appointments? https://www.nuffieldtrust.org.uk/resource/why-does-england-have-unemployed-gps-when-patients-can-t-get-gp-appointments. Accessed: 2026-01-22.

Freedman, S., Golberstein, E., Huang, T.-Y., Satin, D. J., and Smith, L. B. (2021). Docs with their eyes on the

clock? the effect of time pressures on primary care productivity. *Journal of Health Economics*, 77:102442.
Global Government Forum (2026). Nhs app roadmap highlights latest pilots, priorities and plans. https://www.globalgovernmentforum.com/nhs-app-roadmap-highlights-latest-pilots-priorities-and-plans/?utm_source=chatgpt.com. Accessed: 2026-03-17.
Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93:293–300.
Goldman, D. P., Joyce, G. F., Karaca-Mandic, P., et al. (2006). Varying pharmacy benefits with clinical status: the case of cholesterol-lowering therapy. *American Journal of Managed Care*, 12(1):21.
Guan, X., Li, K., and Wu, H. (2025). A blessing or a curse? teletriage service in healthcare. *Management Science*.
Heather, B. (2017). Babylon pilot was not pursued by ccgs. https://www.hsj.co.uk/technology-and-innovation/babylon-pilot-was-not-pursued-by-ccgs/7021131.article. Accessed: 2026-01-14.
Hermanns, B., Kairies-Schwarz, N., Kokot, J., and Vomhof, M. (2025). Heterogeneity in health insurance choice: An experimental investigation of consumer choice and feature preferences. *Journal of Risk and Insurance*, 92(4):1096–1121.
Hoe, T. P. (2022). Does hospital crowding matter? evidence from trauma and orthopedics in england. *American Economic Journal: Economic Policy*, 14(2):231–262.
Iizuka, T. and Shigeoka, H. (2022). Is zero a special price? evidence from child health care. *American Economic Journal: Applied Economics*, 14(4):381–410.
Jakobsson, N. and Svensson, M. (2016). The effect of copayments on primary care utilization: results from a quasi-experiment. *Applied Economics*, 48(39):3752–3762.
Jones, J. (2020). Take it from a paramedic: the ambulance service can't keep up with demand. https://www.theguardian.com/commentisfree/2020/jan/29/paramedic-ambulance-service-call-triage. Accessed: 2026-01-14.
Kairies-Schwarz, N., Kokot, J., Vomhof, M., and Weßling, J. (2017). Health insurance choice and risk preferences under cumulative prospect theory–an experiment. *Journal of Economic Behavior & Organization*, 137:374–397.
Kajaria-Montag, H. and Freeman, M. (2020). Explaining the erosion of relational care continuity: an empirical analysis of primary care in england. *INSEAD Working Paper*.
Kajaria-Montag, H., Freeman, M., and Scholtes, S. (2024). Continuity of care increases physician productivity in primary care. *Management Science*, 70(11):7943–7960.
Kiil, A. and Houlberg, K. (2014). How does copayment for health care services affect demand, health and redistribution? a systematic review of the empirical evidence from 1990 to 2011. *The European Journal of Health Economics*, 15(8):813–828.
Landsem, M. M. and Magnussen, J. (2018). The effect of copayments on the utilization of the gp service in norway. *Social Science & Medicine*, 205:99–106.
Layte, R., Nolan, A., McGee, H., and O'Hanlon, A. (2009). Do consultation charges deter general practitioner use among older people? a natural experiment. *Social Science & Medicine*, 68(8):1432–1438.
Miller, S., Gilbert, S., Virani, V., and Wicks, P. (2020). Patients' utilization and perception of an artificial intelligence–based symptom assessment and advice technology in a british primary care waiting room: exploratory pilot study. *JMIR human factors*, 7(3):e19713.
OECD (2025). *Health at a glance 2025: OECD indicators*. OECD Publishing.
Ostrom, E., Gardner, R., and Walker, J. (1994). *Rules, games, and common-pool resources*. University of Michigan press.
Prolific (2025). Representative samples. https://researcher-help.prolific.com/en/article/95c345. Accessed: 2025-03-31.
Rosen, B., Brammli-Greenberg, S., Gross, R., and Feldman, R. (2011). When co-payments for physician visits can affect supply as well as demand: findings from a natural experiment in israel's national health insurance system. *The International journal of health planning and management*, 26(2):e68–e84.
Schreyögg, J. and Grabka, M. M. (2010). Copayments for ambulatory care in germany: a natural experiment using a difference-in-difference approach. *The European Journal of Health Economics*, 11(3):331–341.
Shurtz, I., Eizenberg, A., Alkalay, A., and Lahad, A. (2022). Physician workload and treatment choice: the case of primary care. *The RAND Journal of Economics*, 53(4):763–791.
Swami, M., Gravelle, H., Scott, A., and Williams, J. (2018). Hours worked by general practitioners and waiting

times for primary care. *Health Economics*, 27(10):1513–1532.

Thami, P. (2025). Truth-telling in a priority pricing mechanism. *Working Paper*.

The Nuffield Trust (2025). General practice staffing tracker. https://www.nuffieldtrust.org.uk/nhs-staffing-tracker/general-practice. Accessed: 2026-01-14.

Vallejo-Torres, L. and Morris, S. (2018). Primary care supply and quality of care in england. *The European Journal of Health Economics*, 19(4):499–519.

Zweifel, P. and Manning, W. G. (2000). Moral hazard and consumer incentives in health care. In *Handbook of Health Economics*, volume 1, pages 409–459. Elsevier.

A Proofs

A.A Proposition 1

Proof. Players i maximizes her utility by choosing either to extract ($x_i = 1$) or not to extract ($x_i = 0$). We assume an overwhelmed system, where own extraction necessarily leads to the exclusion of one of the players.

$$\arg\max_{x_i \in \{0,1\}} \pi_i + \alpha_i \pi_j \mathbb{1}_{z_i=1} - c_i \mathbb{1}_{z_i=1} - c_i^{psych} \mathbb{1}_{Triage \wedge \sigma_{no}}$$

As the action space consists only of two choices, it suffices to compare the resulting utilities to solve this maximization problem. Specifically, player i chooses $x_i = 1$ if and only if $U_i(x_1) > U_i(x_0)$. Thus, we compute the difference of these two utilities:

$$\begin{aligned}\Delta U_i = U_i(x_1) - U_i(x_0) = &\left((1 - \frac{n-1}{n}) p_h(h, \sigma, k)(-A) - \frac{n-1}{n} c_i - c_i^{psych} \mathbb{1}_{Triage \wedge \sigma_{no}} + \alpha_i (\frac{n-1}{n}) h(-A)\right) \\ &- \left(p_h(h, \sigma, k)(-A)\right) \\ \iff \Delta U_i = &\frac{n-1}{n} p_h(h, \sigma, k) A - \frac{n-1}{n} c_i - c_i^{psych} \mathbb{1}_{Triage \wedge \sigma_{no}} - \alpha_i (\frac{n-1}{n}) hA\end{aligned}$$

Thus player i extracts if the following inequality holds:

$$\begin{aligned}&\Delta U_i > 0 \\ \iff &\frac{n-1}{n} p_h(h, \sigma, k) A > \frac{n-1}{n} c_i + c_i^{psych} \mathbb{1}_{Triage \wedge \sigma_{no}} + \alpha_i (\frac{n-1}{n} hA)\end{aligned}$$

□

A.B Corollary 1

Proof. The costs of extraction are independent of k and thus stay constant with a change in k. The benefit of extraction after receiving a no harm signal is:

$$\begin{aligned}Benefit_i &= \frac{n-1}{n} p_{\text{high}}(\theta_h | \sigma_{no}) A \\ \iff Benefit_i &= \frac{n-1}{n} \frac{(1-k)h}{(1-k)h + k(1-h)} A\end{aligned}$$

Taking the derivative w.r.t. k:

$$\frac{\partial Benefit_i}{\partial k} = -\frac{(n-1)h(1-h)A}{n(2hk - h - k)^2} < 0$$

Thus, player i is more likely to switch to no extraction – given that they extract in the first place. If player i already does not extract in a no-harm signal scenario, she would not switch to extraction if the no-harm signal became more accurate. Thus, increasing the signal accuracy k leads to a weak decrease in extraction choices. □

A.C Corollary 2

Proof. The benefits of extraction are independent of c_i and thus stay constant with a change in c_i. The costs of extraction increase with the rate $\frac{\partial Costs_i}{\partial c_i} = \frac{n-1}{n} > 0$ and thus extraction choices (weakly) decrease after the introduction of an extraction cost c. □

A.D Corollary 3

Proof. The introduction of psychic costs does not change the benefit of extraction. Assuming that α_i remains constant, an increase in $c_i^{psych}\mathbb{1}_{Triage \wedge \sigma_{no}}$ leads to a (weak) decrease in extraction choices after receiving a no harm signal as $\frac{\partial Costs_i}{\partial c_i^{psych}\mathbb{1}_{Triage \wedge \sigma_{no}}} = 1 > 0$. □

A.E Corollary 4

Proof. Absent psychic moral costs, the inequality of Proposition 1 reduces to:

$$\frac{n-1}{n} p_h(h, \sigma, k)A > \frac{n-1}{n} c_i + \alpha_i \frac{n-1}{n} hA$$
$$\iff (p_h(\theta_h|\sigma_h) - \alpha_i h)A > c_i$$

Note that $p_h(\theta_h|\sigma_h) > h > \alpha_i h$ for $\forall\ k \in (0.5, 1)$ and $\forall\ \alpha_i \in [0, 1)$. Absent extraction costs c_i, player i would always extract.

□

B Experiment 1

B.A Pre-registration

ASPREDICTED

Wharton CREDIBILITY LAB
University of Pennsylvania

Institutional factors and the extraction of limited health resources (#205066)

Author(s)
Diya Abraham (University of Reading) - d.e.abraham@reading.ac.uk
Ondřej Krčál (Masaryk University) - ondrej.krcal@econ.muni.cz
Jonathan Stäbler (University of Mannheim) - jonathanstaebler@gmail.com

Pre-registered on:
2024/12/18 01:41 (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
The healthcare situation in many developed countries has become precarious due to limited supply and a steadily growing demand stemming from mandatory health cover for all. As a result, health resources may get allocated inefficiently with those with potentially serious conditions not getting treated in a timely manner. Our research question concerns the effectiveness of measures aimed at preventing the inefficient allocation of limited health resources. These measures include: i) instituting a small cost of drawing on health services, ii) creating a triage system based on self-reporting one's symptoms that introduces psychological costs of reporting a type that is contrary to one's signal and iii) receiving access to more accurate signals about one's health needs.

3) Describe the key dependent variable(s) specifying how they will be measured.
We model limited supply of resources in the healthcare system using a variation of a one-shot common pool resource game with heterogeneous types, namely, high-harm types (probability h) or no-harm types (probability 1-h), each of whom receive private signals, sig_h = 0 or 1 (where sig_h = 1 if they receive the signal they are a high-harm type), of accuracy k about their harm type. Participants are divided into groups of 4 and start with the same initial endowment. Losses (A) are incurred by high-harm types if they do not extract from the common resource pool available to their group. No-harm types incur no losses. Before learning their type, each member of the group is asked to make a binary decision (q_i = 0 or 1) of whether to extract resources from the common resource pool given h, k and sig_h.

Since resources are limited, only 3 of the 4 members of the group can extract resources from the common pool at a time. If all 4 members decide to extract, a random draw will determine which 3 members are allowed to extract and thus offset their losses if they end up being high-harm types. If 3 or fewer members decide to extract, no group member that extracts will incur any loss. A high-harm type can thus incur the losses in two situations: i) if they do not extract and ii) if they are not selected to extract in the random draw. We elicit participants' extraction decisions in strategy method where they provide their unconditional extraction decisions for sig_h = 0 and sig_h = 1 as well as their conditional extraction decisions, where the conditionality is generated by (i) the signal they receive sig_h (= 0 or 1) and (ii) the number of other group members who chose to extract, Q_j, where Q_j = 1 if all 3 of their group members chose to extract and Q_j = 0 if 2 or fewer of their group members chose to extract.

We assume players have other-regarding preferences represented by a parameter alpha. We then arrive at the pure-strategy Nash equilibrium in which all players extract if the following condition holds: p(h, sig_h, k) > alpha*h, where p is the agent's belief (we assume Bayesian updating) that they are a high-harm type knowing h and after receiving sig_h of accuracy k. If, for at least one player, this condition does not hold, i.e., p(h, sig_h, k) < alpha*h, then there exists an asymmetric pure-strategy Nash equilibrium, where just this player does not choose to extract while the rest do. In this study, we are interested in this asymmetric pure-strategy equilibrium, where all but one player extracts.

Our main dependent variable, ExtractionCond, is the participant's extraction decision when sig_h = 0 and Q_j = 1.

4) How many and which conditions will participants be assigned to?
Participants will be randomly assigned to 4 between-subject treatments, namely, Baseline, Cost, Triage and Accuracy. All participants receive the same fixed payment. Each of them also start with an additional initial endowment of GBP 2.5 (their bonus payment). The loss A is equal to their bonus payment. The probability of being a high-harm, i.e. h, is 50%. The likelihood that the signal, k, they receive about their type is true is 66.67%.

The Baseline with the above parameters is exactly as described above in pt. 3.

In the Cost treatment, all parameters remain the same but the participants who get to extract will incur an additional extraction cost of GBP 0.50 from their bonus payment.

In the Triage treatment, the parameters are the same as in the Baseline but the extraction decision is framed as a reporting decision, where rather than deciding whether to extract, participants are asked to report whether they are a high-harm or no-harm type with only those who report they are high-harm types being considered for extraction. In this treatment, reporting that one is a high-harm type is therefore equivalent to choosing to extract in the other treatments.

In the Accuracy treatment, the only difference relative to the Baseline is that the parameter k (i.e., the accuracy of the signal) is 90%.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We derive the following theoretical predictions based on our model:
Prediction 1: ExtractionCond_Base > ExtractionCond_Cost
Prediction 2: ExtractionCond_Base > ExtractionCond_Triage

Prediction 3: ExtractionCond_Base > ExtractionCond_Accuracy

We will use a one-sided proportion to test the above predictions. As a robustness check, we will run Probit models regressing ExtractionCond on the relevant treatment dummies along with demographic controls.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We exclude observations that did not complete the experiment and observations with Prolific IDs that are not listed by Prolific as having completed the experiment.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect at least 1560 participants in total, with 390 in each of the 4 treatments.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We explore whether empirical expectations, normative expectations and personal norms around non-extraction vary across our treatments.
As further exploratory analyses, we will re-run our main analysis after excluding those who did not extract after receiving a high-signal in all three of these decisions.

B.B Robustness Checks Results

In this section, we present three robustness checks of which the first two were pre-registered. In the first robustness check, we run a Probit model, where we control for individual characteristics of the participants (see Table 7). In robustness check 2, we exclude participants, who never extract after receiving a high-harm signal, even when they know that not all others are extracting. Robustness check 3 excludes those participants, who have at least 3 wrong answers to the quiz questions. In robustness checks 2 & 3, we use two-sided proportion tests as for the main comparisons. In all three robustness checks, all pairwise comparisons between the three treatments with the Baseline remain statistically significant (at least $p < 0.01$).

TABLE 7: Probit regression of extraction choices after a no-harm signal conditional on all others extracting

	Dependent variable:
	Extraction no-harm all others extract
Treatment: Cost	-0.45^{***}
	(0.09)
Treatment: Triage	-0.59^{***}
	(0.09)
Treatment: Accuracy	-0.26^{**}
	(0.09)
Female	-0.12
	(0.06)
Age: 31-40	0.03
	(0.10)
Age: 41-50	0.11
	(0.11)
Age: 51-60	0.001
	(0.10)
Age: 61+	-0.04
	(0.10)
Ethnicity: Black	0.01
	(0.12)
Ethnicity: Asian	-0.10
	(0.18)
Ethnicity: Other	-0.03
	(0.17)
Constant	-0.02
	(0.10)
Observations	1,696
Log Likelihood	$-1,072.66$
Akaike Inf. Crit.	2,169.31

*Note: Robust (HC1) standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001*

B.C No-Harm Signal Heterogeneity

TABLE 8: Probit Regression with Interaction Terms

	Dependent variable:		
	Extraction no-harm all others extract		
	(1)	(2)	(3)
Cost	-0.56^{***}	-0.47^{***}	-0.38
	(0.13)	(0.13)	(0.22)
Triage	-0.52^{***}	-0.63^{***}	-0.46^{*}
	(0.13)	(0.13)	(0.22)
Accuracy	-0.29^{*}	-0.33^{**}	-0.19
	(0.12)	(0.12)	(0.23)
Female	-0.15		
	(0.12)		
Age > 48		-0.10	
		(0.12)	
White			0.09
			(0.16)
Cost × Female	0.21		
	(0.18)		
Triage × Female	-0.14		
	(0.18)		
Accuracy × Female	0.06		
	(0.17)		
Cost × Age > 48		0.05	
		(0.18)	
Triage × Age > 48		0.09	
		(0.18)	
Accuracy × Age > 48		0.16	
		(0.17)	
Cost × White			-0.09
			(0.24)
Triage × White			-0.15
			(0.25)
Accuracy × White			-0.08
			(0.25)
Constant	0.01	-0.02	-0.14
	(0.09)	(0.08)	(0.15)
Observations	1,696	1,696	1,696
Log Likelihood	$-1{,}071.99$	$-1{,}075.01$	$-1{,}075.35$
Akaike Inf. Crit.	2,159.97	2,166.01	2,166.70

*Note: Robust (HC1) standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001*

B.D Choices per Signal and Condition

TABLE 9: Probit regression extraction choices per signal and condition

	Dependent variable:
	Extraction
No-harm Signal	-0.71^{***}
	(0.04)
Not All Others Extract	0.28^{***}
	(0.04)
Unconditional	0.22^{***}
	(0.03)
No-harm Signal: Not All Others Extract	0.29^{***}
	(0.05)
No-harm Signal: Unconditional	0.07
	(0.04)
Constant	0.33^{***}
	(0.03)
Observations	10,176
# Clusters	1,696
Log Likelihood	−6,560.24
Akaike Inf. Crit.	13,132.47

*Note: SE clustered at individual level. * p<0.05, ** p<0.01, *** p<0.001*

B.E Efficiency

For each individual, we construct an efficiency measure by computing the expected probability that a high-harm player is excluded given the individual's extraction choices after receiving a no-harm and a high-harm signal (conditional on all others extracting). If player i extracts, one of the four players is randomly excluded. Whether the excluded player is high-harm depends on the prior probability h for the remaining group members and on the posterior probability that player i is high-harm conditional on the received signal. If player i does not extract, the exclusion probability is simply the posterior probability that player i is high-harm. The overall individual measure takes the weighted averages over the resulting expected exclusion probabilities for all possible harm distributions, using the actual extraction choices of player i depending on the signal, where each signal occurs with probability h.

Figure 6 depicts the average expected efficiency per treatment. The red dotted lines represent the upper and lower bounds of decision efficiency. The maximum decision efficiency assumes that no one extracts after receiving a no-harm signal and everyone extracts after a high-harm signal. The minimum decision efficiency assumes that everyone extracts after receiving a no-harm signal and none extracts after receiving a high-harm

signal. Note that the maximum decision efficiency is not zero because players do not know their true harm type but rather rely only on a noisy signal about their harm type. Consequently, if all those who receive a no-harm signal refrain from extracting, there will be some high-harm players among them simply because the signal is incorrect. Similarly, a high-harm signal receiver can end up being of no-harm and take the spot away from an actual no-harm player. Such 'mistakes' are less frequent in the *Accuracy* treatment, where they occur in only 10% of cases (compared to 33.3% of cases in the other treatments). Thus the minimum and maximum decision efficiencies are broader in the *Accuracy* treatment (as seen in Figure 6).

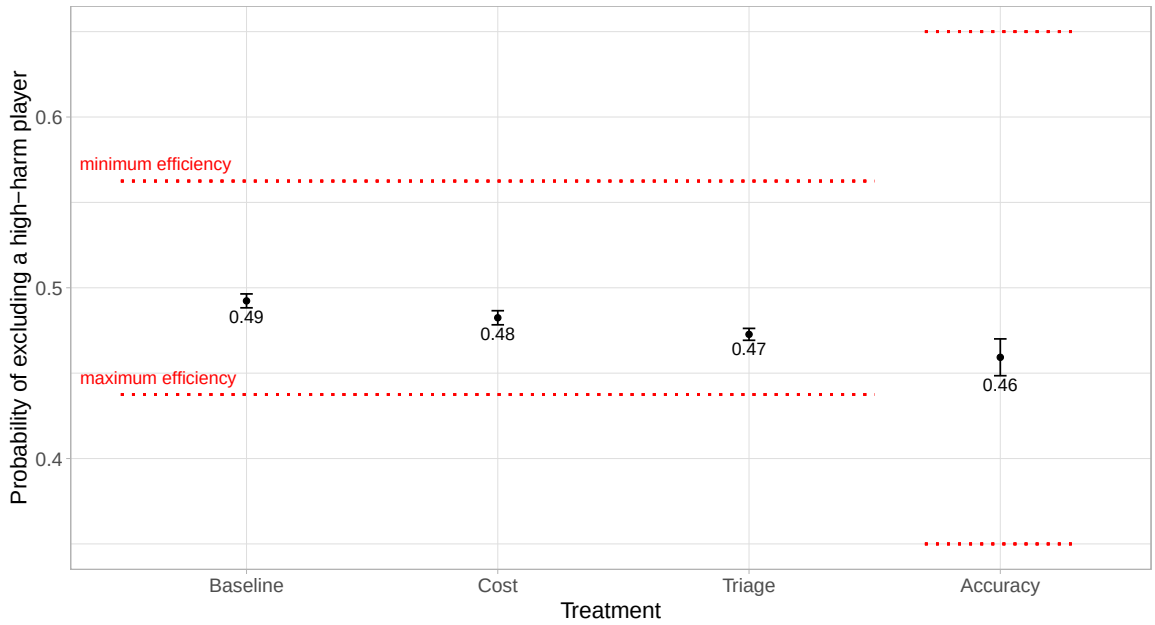


FIGURE 6: EXTRACTION EFFICIENCY

Figure 6 shows a clear ordering in terms of efficiency. The expected probability of excluding a high-harm player is highest in the Baseline (0.49), followed by the *Cost* treatment (0.48), the *Triage* treatment (0.47), and is lowest in the *Accuracy* treatment (0.46). All pairwise comparisons are significant (at least $p < 0.01$, two-sided Wilcoxon rank sum tests).

While these differences may appear modest in absolute terms, even a one percentage point reduction in the probability that a high-harm player is excluded can correspond to substantial welfare gains when the consequences of untreated high-harm conditions are severe.

Moreover, the overall differences are likely attenuated by potential noise in extraction choices after receiving high-harm signals. We note here that since the main focus of Experiment 1 was to identify interventions that reduce low-priority demand in an overwhelmed system, the instructions and comprehension questions were specifically designed to ensure participants understood the incentives associated with receiving a no-harm signal when all other players in one's group had chosen extract. As such, a certain amount of non-extraction after receiving a high-harm signal is not unexpected.

To isolate the direct efficiency effects of more accurate information for low-priority individuals, Figure 7 additionally reports efficiency conditional on receiving a no-harm signal only. In this case, the *Accuracy* treatment reduces the probability that a high-harm player is excluded by 18 %-points relative to the Baseline. This highlights the substantial efficiency gains from providing more accurate information to those receiving a no-harm signal.

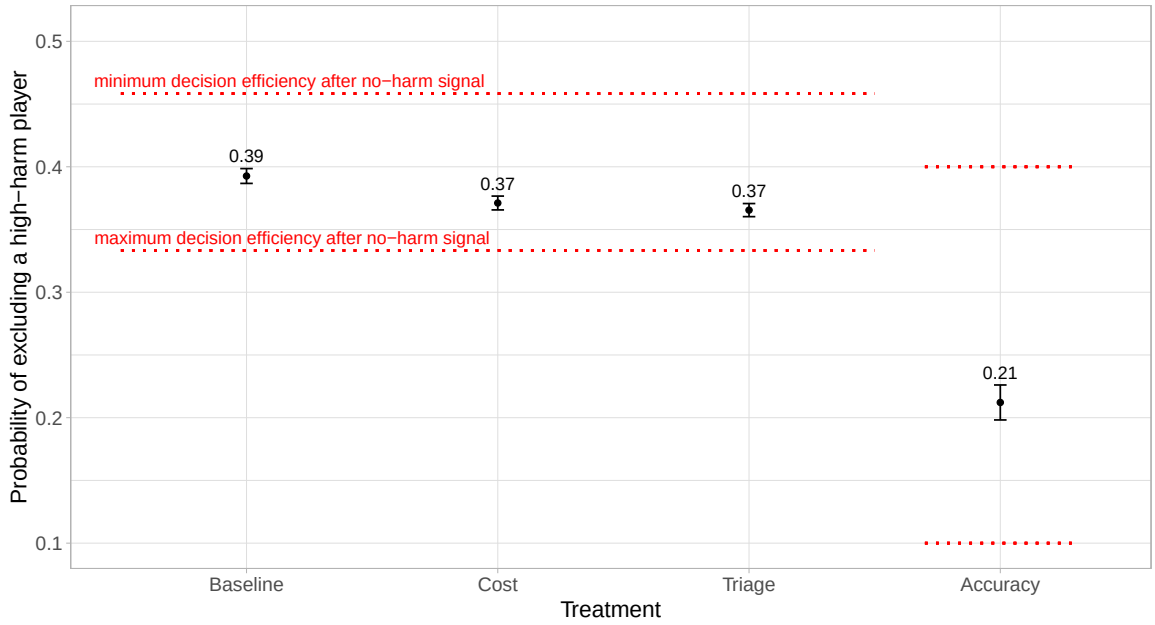
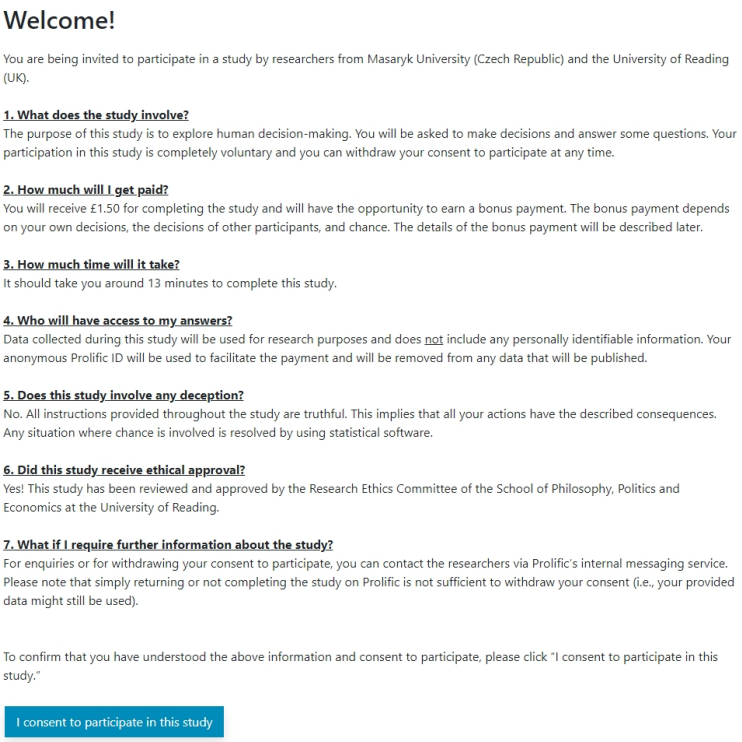


FIGURE 7: EXTRACTION EFFICIENCY AFTER RECEIVING A NO-HARM SIGNAL

B.F Instructions

This section presents screenshots of the environment used in Experiment 1, in the order in which they were shown to participants. All screens correspond to the *Baseline* treatment, except for Screen 12, which is shown in four treatment-specific variants. Screens 13–17 vary across treatments in the same manner as Screen 12. Other screens may also differ but only marginally across treatments, but these differences are not displayed here. The full set of instructions is available upon request.



Welcome!

You are being invited to participate in a study by researchers from Masaryk University (Czech Republic) and the University of Reading (UK).

1. What does the study involve?
The purpose of this study is to explore human decision-making. You will be asked to make decisions and answer some questions. Your participation in this study is completely voluntary and you can withdraw your consent to participate at any time.

2. How much will I get paid?
You will receive £1.50 for completing the study and will have the opportunity to earn a bonus payment. The bonus payment depends on your own decisions, the decisions of other participants, and chance. The details of the bonus payment will be described later.

3. How much time will it take?
It should take you around 13 minutes to complete this study.

4. Who will have access to my answers?
Data collected during this study will be used for research purposes and does not include any personally identifiable information. Your anonymous Prolific ID will be used to facilitate the payment and will be removed from any data that will be published.

5. Does this study involve any deception?
No. All instructions provided throughout the study are truthful. This implies that all your actions have the described consequences. Any situation where chance is involved is resolved by using statistical software.

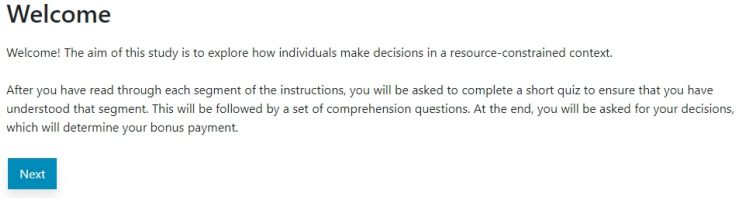
6. Did this study receive ethical approval?
Yes! This study has been reviewed and approved by the Research Ethics Committee of the School of Philosophy, Politics and Economics at the University of Reading.

7. What if I require further information about the study?
For enquiries or for withdrawing your consent to participate, you can contact the researchers via Prolific's internal messaging service. Please note that simply returning or not completing the study on Prolific is not sufficient to withdraw your consent (i.e., your provided data might still be used).

To confirm that you have understood the above information and consent to participate, please click "I consent to participate in this study."

I consent to participate in this study

Screen 1



Welcome

Welcome! The aim of this study is to explore how individuals make decisions in a resource-constrained context.

After you have read through each segment of the instructions, you will be asked to complete a short quiz to ensure that you have understood that segment. This will be followed by a set of comprehension questions. At the end, you will be asked for your decisions, which will determine your bonus payment.

Next

Screen 2

Instructions

Payment

In addition to your fixed payment for this study, you can earn a bonus of up to £2.50 depending on your decisions, that of other participants in this study, and chance.

Your Type

When the study begins, every participant will be assigned a "type" and will be **endowed with the full bonus payment of £2.50**. There are two possible types: **high-harm types** and **no-harm types**. Your type is decided by a virtual coin flip (50-50 chance of either type). If the coin lands on heads, you will be a high-harm type, and if it lands on tails, you will be a no-harm type. Every participant is thus **equally likely to be a high-harm or a no-harm type**.

If you are a **high-harm type**, you **stand to lose your bonus payment of £2.50**. If you are a **no-harm type**, you will incur **no losses and keep your £2.50 bonus payment**.

Your Group

After the study, all participants will be randomly and anonymously assigned to groups of 4 people. Your decisions and the decisions of the other 3 members of your group will be used to determine the bonus payment for each of you.

Quiz Questions

Q1: What is the chance of being a no-harm type?

Q2: How many people are there in each group (including you)?

Q3: What are the potential losses no-harm types can incur?

Q4: What are the potential losses high-harm types can incur?

Next

Screen 3

Quiz Questions: Correct Answers

You have correctly answered all questions!

Q1: What is the chance of being a no-harm type?
Correct: The chance of being a no-harm type is 50%.

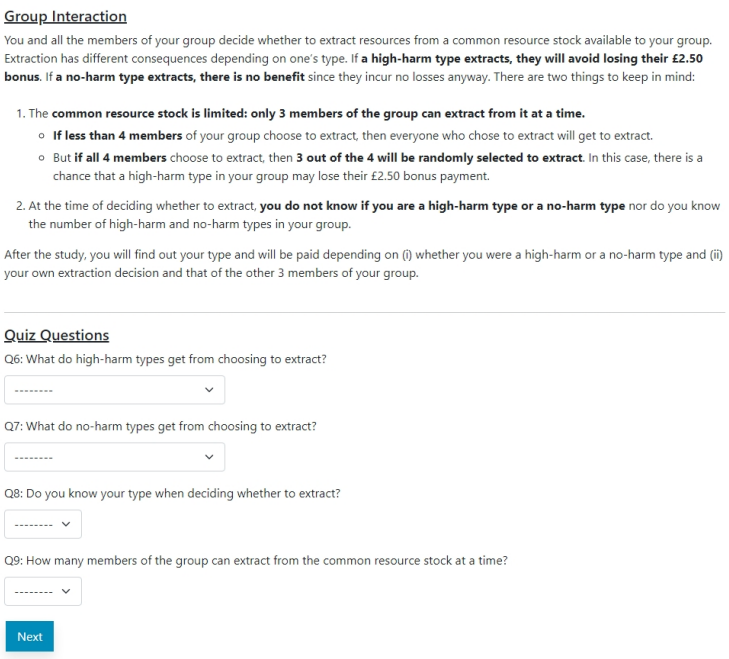
Q2: How many people are there in each group (including you)?
Correct: Every participant is randomly and anonymously assigned to a group of 4 people.

Q3: What are the potential losses no-harm types can incur?
Correct: No-harm types incur no losses and so get to keep their £2.50 bonus.

Q4: What are the potential losses high-harm types can incur?
Correct: High-harm types may lose their bonus payment of £2.50.

Next

Screen 4



Group Interaction

You and all the members of your group decide whether to extract resources from a common resource stock available to your group. Extraction has different consequences depending on one's type. If **a high-harm type extracts, they will avoid losing their £2.50 bonus**. If **a no-harm type extracts, there is no benefit** since they incur no losses anyway. There are two things to keep in mind:

1. The **common resource stock is limited: only 3 members of the group can extract from it at a time.**
   - **If less than 4 members** of your group choose to extract, then everyone who chose to extract will get to extract.
   - But **if all 4 members** choose to extract, then **3 out of the 4 will be randomly selected to extract**. In this case, there is a chance that a high-harm type in your group may lose their £2.50 bonus payment.
2. At the time of deciding whether to extract, **you do not know if you are a high-harm type or a no-harm type** nor do you know the number of high-harm and no-harm types in your group.

After the study, you will find out your type and will be paid depending on (i) whether you were a high-harm or a no-harm type and (ii) your own extraction decision and that of the other 3 members of your group.

Quiz Questions

Q6: What do high-harm types get from choosing to extract?

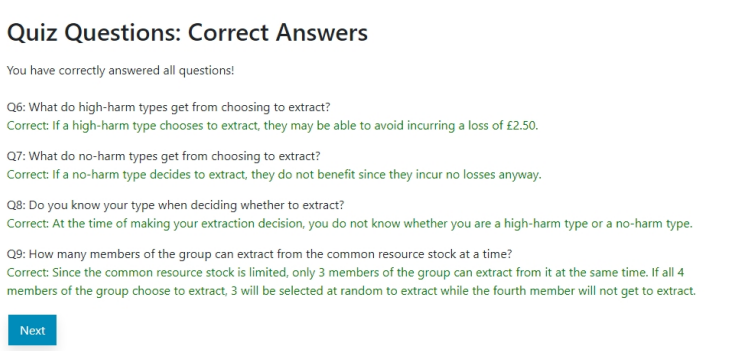
Q7: What do no-harm types get from choosing to extract?

Q8: Do you know your type when deciding whether to extract?

Q9: How many members of the group can extract from the common resource stock at a time?

Next

SCREEN 5



Quiz Questions: Correct Answers

You have correctly answered all questions!

Q6: What do high-harm types get from choosing to extract?
Correct: If a high-harm type chooses to extract, they may be able to avoid incurring a loss of £2.50.

Q7: What do no-harm types get from choosing to extract?
Correct: If a no-harm type decides to extract, they do not benefit since they incur no losses anyway.

Q8: Do you know your type when deciding whether to extract?
Correct: At the time of making your extraction decision, you do not know whether you are a high-harm type or a no-harm type.

Q9: How many members of the group can extract from the common resource stock at a time?
Correct: Since the common resource stock is limited, only 3 members of the group can extract from it at the same time. If all 4 members of the group choose to extract, 3 will be selected at random to extract while the fourth member will not get to extract.

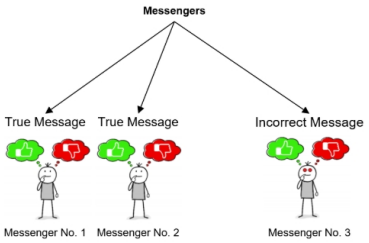
Next

SCREEN 6

Your Message

Before deciding whether to extract, you will receive **private information** in the form of a message about whether you are a high-harm type or a no-harm type. You will receive either a positive message "*You are a no-harm type and so incur no loss to your £2.50 bonus payment*" or a negative message "*You are a high-harm type and so may lose your £2.50 bonus payment*".

The messages are provided by one of the 3 messengers depicted below. But note that not all of these messengers are trustworthy. While 2 of them are trustworthy and *always* tell the truth, 1 of them is untrustworthy and will always present you with an incorrect message about your type. The computer will randomly select one of these 3 messengers to deliver the message, **but you will not find out which messenger (trustworthy or untrustworthy) the computer selected**.



Since the computer randomly selects one of them to deliver your message, you will receive a true message with two-thirds (67%) chance and an incorrect message with one-third (33%) chance. This means that if, for example, you receive a message telling you that you are a no-harm type, there is a one-third (33%) chance that you are really a high-harm type.

Quiz Questions

Q10: What is the chance that you receive a true message about your type?

Q11: What is the chance that you receive an incorrect message about your type?

Next

SCREEN 7

Quiz Questions: Correct Answers

You have correctly answered all questions!

Q10: What is the chance that you receive an accurate message about your type?
Correct: There is a 67% chance that the message you receive is true.

Q11: What is the chance that you receive an incorrect message about your type?
Correct: There is a 33% chance that the message you receive is incorrect.

Next

SCREEN 8

Comprehension Questions

Before you begin, please answer all the questions for the following situation. You can also review the instructions by clicking on the button "show" at the bottom of the page.

Imagine you have just received the following message:
"*You are a no-harm type and so incur no loss to your £2.50 bonus payment*".
Remember: There is a 67% chance that this message is true and a 33% chance that it is incorrect.

Further, suppose you know that all the other 3 members of your group have decided to extract.

Q12: Given this message, what is the chance that you end up being a high-harm type?

Now imagine that you have also chosen to extract.

Q13: What will your bonus payment be if you end up being a **no-harm type** and **are** selected to extract?

Q14: What will your bonus payment be if you end up being a **high-harm type** and **are** selected to extract?

Q15: What will your bonus payment be if you end up being a **high-harm type** and are **not** selected to extract?

Now imagine that you have chosen **not** to extract.

Q16: What will your bonus payment be if you end up being a **high-harm type**?

Q17: What will your bonus payment be if you end up being a **no-harm type**?

Q18: How many of the others will get to extract?

Next

Overview of Instructions ↓ show ↓

Overview of Instructions ↑ hide ↑

Your Type

When the study begins, every participant will be assigned a "type" and will be endowed with the full bonus payment of £2.50. You can be a high-harm type or a no-harm type, each with equal likelihood (50-50 chance). High-harm types stand to lose their bonus payment of £2.50 if they do not extract resources from a common resource stock. No-harm types incur no loss of their bonus payment.

Group Interaction

After the study, participants will be randomly and anonymously assigned to groups of 4. During the study, all members of the group decide whether to extract resources from the common resource stock. If a high-harm type extracts, they will avoid losing their £2.50 bonus. If a no-harm type extracts, there is no benefit since they incur no losses anyway.

There are two things to keep in mind:

1. The **common resource stock is limited: only 3 members of the group can extract from it at a time.**
   - If **less than 4 members** of your group choose to extract, then everyone who chose to extract will get to extract.
   - But **if all 4 members** choose to extract, then **3 out of the 4 will be randomly selected to extract**. In this case, there is a chance that a high-harm type in your group may lose their £2.50 bonus payment.
2. At the time of deciding whether to extract, **you do not know if you are a high-harm type or a no-harm type** nor do you know the number of high-harm and no-harm types in your group.

After the study, you will find out your type and will be paid depending on:

- Whether you were a high-harm or a no-harm type.
- Your own extraction decision and that of the other 3 members of your group.

Your Message

Before making your choice, you will receive a message about your type. Your message always has a two-thirds (67%) chance of being true and thus a one-third (33%) chance of being incorrect.

SCREEN 9

Quiz Questions: Correct Answers

You have correctly answered all questions!

Imagine you have just received the following message:
"*You are a no-harm type and so incur no loss to your £2.50 bonus payment*".
Remember: There is a 67% chance that this message is true and a 33% chance that it is incorrect.

Further, suppose you know that all the other 3 members of your group have decided to extract.

Q12: Given this message, what is the chance that you end up being a high-harm type?
Correct: Remember that the message that you received is incorrect with 33% chance. In this scenario you have received the message: *You are a no-harm type and so incur no loss to your £2.50 bonus payment*, hence you have a 33% chance that you actually end up being a high-harm type (and 67% chance that you end up being a no-harm type).

Imagine that you have also chosen to extract.

Q13: What will your bonus payment be if you end up being a **no-harm type** and **are** selected to extract?
Correct: Since you are a no-harm type, you incur no losses. Thus, your net bonus payment is £2.50.

Q14: What will your bonus payment be if you end up being a **high-harm type** and **are** selected to extract?
Correct: Since you are high-harm type who gets to extract, you will avoid losing your bonus payment of £2.50. This brings your net bonus payment to £2.50.

Q15: What will your bonus payment be if you end up being a **high-harm type** and are **not** selected to extract?
Correct: Since you are high-harm type who does not get to extract, you will lose your bonus payment of £2.50. This brings your net bonus payment to £0.

Now imagine that you have chosen **not** to extract.

Q16: What will your bonus payment be if you end up being a **high-harm type**?
Correct: Since you are high-harm type who does not extract, you will lose your bonus payment of £2.50. This brings your net bonus payment to £0.

Q17: What will your bonus payment be if you end up being a **no-harm type**?
Correct: Since you are a no-harm type, you incur no losses. This brings your net bonus payment to £2.50.

Q18: How many of the others will get to extract?
Correct: Since you decided not to extract, all others will be able to extract.

Next

Screen 10

Decision Scenarios

Your type as well as the message you receive have already been determined.

On the next screens, we will ask you to make **6 decisions depending on several possible situations you might find yourself in**. All these decisions pertain to the same group interaction explained previously.

Any of the 6 scenarios presented to you on the next screens might be the one you are actually in and thus the one that matters for your payment. You should thus pay close attention to the features of each scenario while making your decisions.

Next

Click "show" to see an overview of the six decisions and how one of them is randomly selected to be paid.

Why is this decision relevant for my bonus payment? ↑ hide ↑

You make 6 decisions in total.

Decisions 1, 2 and 3 correspond to a situation where you have received the message you are a *no*-harm type and Decisions 4, 5 and 6 correspond to a situation where you have received the message that you are a *high*-harm type. Note that your actual message has already been determined at this time but since you do not know what it is, you should make your decisions in both sets of scenarios as though that is the message you have actually received.

Decision Scenarios 1, 2 and 3 are as follows:

- *Decision 1:* What do you choose to do if you have received the message that you are a no-harm type and all the three other members of your group have decided to extract? (extract/ not extract)
- *Decision 2:* What do you choose to do if you have received the message that you are a no-harm type and fewer than three other members of your group have decided to extract? (extract/ not extract)
- *Decision 3:* What do you choose to do if you have received the message that you are a no-harm type independent of what the three other members of your group have decided to do? (extract/ not extract)

We refer to Decisions 1 and 2 as your *conditional* decisions (as they are *conditional* on what your group members do) and Decision 3 as your *unconditional* decision (as it is independent of what your group members do). Decisions 4, 5 and 6 are exactly the same as Decisions 1, 2 and 3 except this time you will be making decisions for the situation where you have received the message that you are a *high*-harm type. As before, we will refer to Decisions 4 and 5 as your *conditional* decisions and Decisions 6 as your *unconditional* decision.

To determine which of these 6 decisions (two unconditional and four conditional decisions) matters for a given participant's payment, for each group, 3 members of the group are randomly selected and these 3 members are paid according to their unconditional extraction decision and the *actual* message they received. The remaining member of the group is paid according to one of their four conditional decisions. The exact conditional decision that is relevant for this member's payment depends on:

(i) the unconditional decisions of the other 3 members of their group (i.e., whether all three or fewer than three have chosen to extract) and

(ii) the actual message the last member received.

Since you do not know which message you actually received nor whether your conditional or unconditional decisions will matter for your payment, you should make your decisions in all six decision scenarios as though that will be the one that counts for your payment.

Screen 11

Decision 1

Now you are in the following situation:
[Click the button 'show' below to see why this situation is relevant for your payment]

(i) **All the three other members of your group have decided to extract.**

(ii) You have received the following message:

*"You are a **no-harm type** and so incur no loss to your £2.50 bonus payment"*
Remember: There is a 67% chance that this message is true and a 33% chance that it is incorrect.

What would you like to do in this situation?

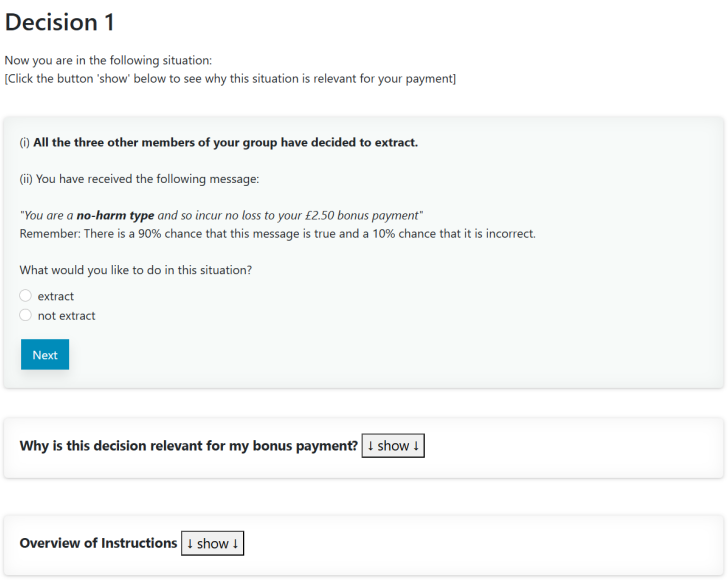
- ○ extract
- ○ not extract

Next

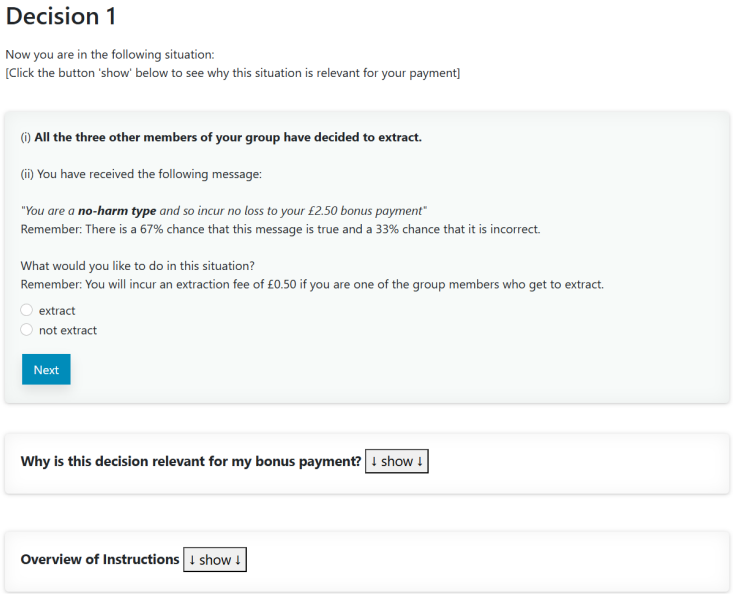
Why is this decision relevant for my bonus payment? ↓ show ↓

Overview of Instructions ↓ show ↓

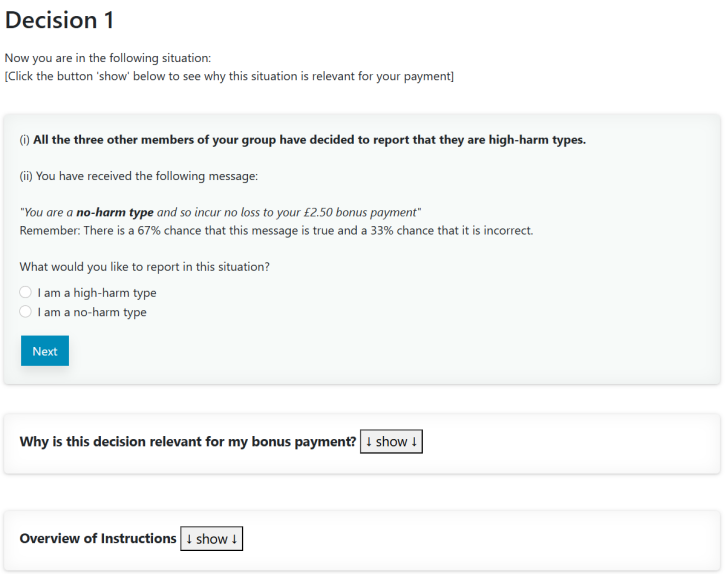
Screen 12 – Baseline



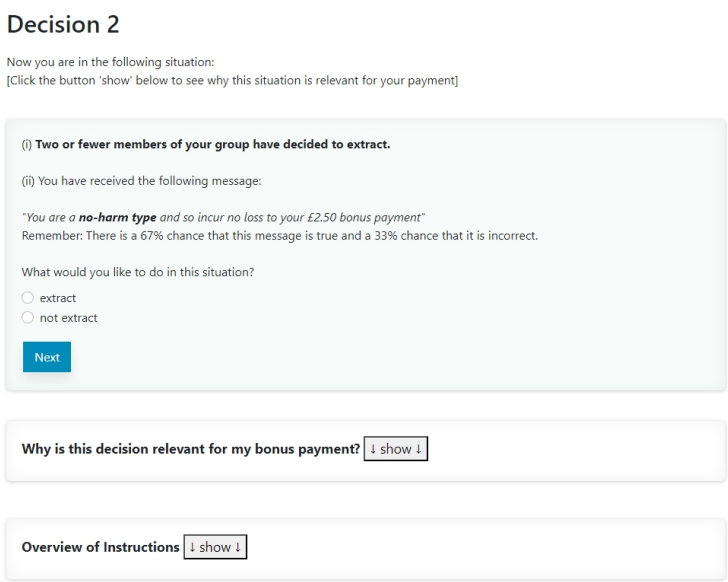
Screen 12 – Accuracy



Screen 12 – Cost



Screen 12 – Triage



Screen 13

Decision 3

Now you are in the following situation:
[Click the button 'show' below to see why this situation is relevant for your payment]

You have received the following message:

*"You are a **no-harm type** and so incur no loss to your £2.50 bonus payment"*
Remember: There is a 67% chance that this message is true and a 33% chance that it is incorrect.

What would you like to do in this situation (independent of what your other group members decide to do)?

- ○ extract
- ○ not extract

Next

Why is this decision relevant for my bonus payment? ↓ show ↓

Overview of Instructions ↓ show ↓

Screen 14

Decision 4

Now you are in the following situation:
[Click the button 'show' below to see why this situation is relevant for your payment]

(i) **All the three other members of your group have decided to extract.**

(ii) You have received the following message:

*"You are a **high-harm type** and so may lose your £2.50 bonus payment"*
Remember: There is a 67% chance that this message is true and a 33% chance that it is incorrect.

What would you like to do in this situation?

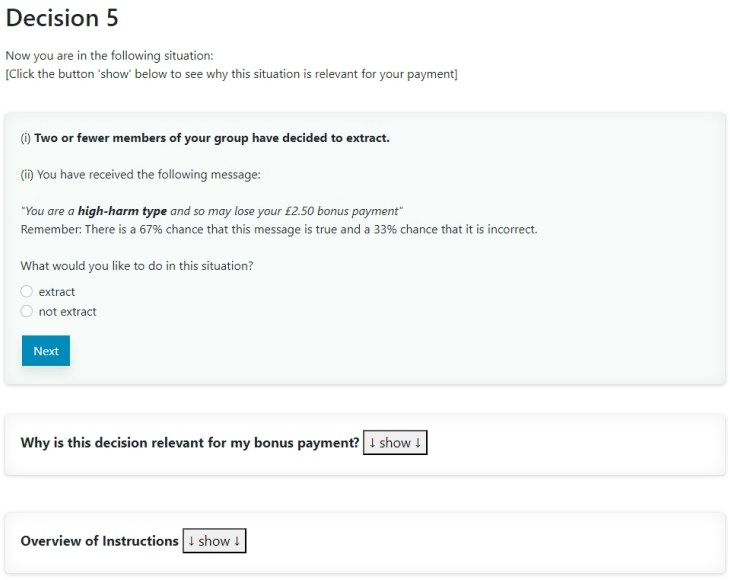
- ○ extract
- ○ not extract

Next

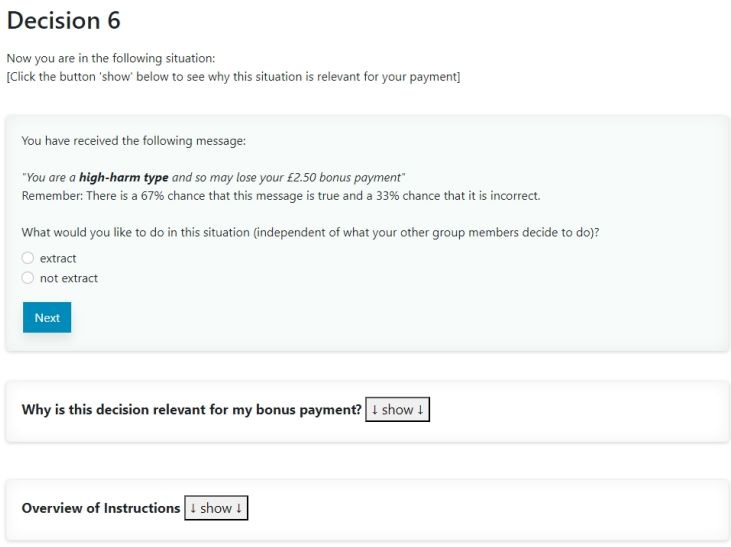
Why is this decision relevant for my bonus payment? ↓ show ↓

Overview of Instructions ↓ show ↓

Screen 15



Screen 16



Screen 17

Questionnaire

All your decisions are now completed. We are interested in what you think about the situation you faced in Decision 1. In this situation:

(i) **All the three other members of your group have decided to extract.**

(ii) You have received the following message:

*"You are a **no-harm type** and so incur no loss to your £2.50 bonus payment"*
Remember: There is a 67% chance that this message is true and a 33% chance that it is incorrect.

What would you like to do in this situation?

How much do you agree with the following statements?
[0 = completely disagree, 100 = completely agree]

Click on the blue bars to start using the sliders to enter a value.

Q1: Most people **do not** choose to extract in the above situation.

0 — 100

Completely disagree — Your choice: **38** — *Completely agree*

Q2: Most people **should not** choose to extract in the above situation.

0 — 100

Completely disagree — *Completely agree*

Q3: Most others think that people **should not** choose to extract in the above situation.

0 — 100

Completely disagree — *Completely agree*

Submit

Screen 18

Final questionnaire

Please answer some final questions below and click "Submit" to complete the study.

Q1: What is your age?

Q2: What is your gender?

Q3: What best describes your employment status?

If your answer is other, please specify:

Q4: Is Prolific (maybe together with other online survey or crowd working platforms) your primary source of income?

Q5: What is your highest level of education?

If your answer is other, please specify:

Submit

Screen 19

C Experiment 2

C.A Pre-registration

ASPREDICTED

Wharton CREDIBILITY LAB
University of Pennsylvania

Institutional factors and the extraction of limited health resources: follow-up (#252763)

Author(s)
Diya Abraham (University of Reading) - d.e.abraham@reading.ac.uk
Ondřej Krčál (Masaryk University) - ondrej.krcal@econ.muni.cz
Jonathan Stäbler (Masaryk University) - jonathanstaebler@gmail.com

Pre-registered on:
2025/10/20 02:04 (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
We study how individuals make decisions about their health in 5 hypothetical scenarios, 4 of which describe symptoms that do not require medical attention, and 1 of which describes symptoms that does require medical attention. These 5 scenarios were created in consultation with practicing medical professionals in the UK and were further vetted by 14 online medical professionals based in the UK. In our study, participants read each scenario and state how likely they are to book an appointment with their General Practitioner (GP). We test whether the chance of an unnecessary GP visit (i.e., in a scenario that does not warrant medical attention) can be reduced by introducing these measures: 1) instituting a small upfront cost that is incurred by the patient when visiting their GP, 2) providing more information about their current health condition through active reflection or speaking to a friend in the medical profession, and 3) implementing a system whereby they would need to exaggerate their symptoms in their GP's online triage form to secure the appointment.

3) Describe the key dependent variable(s) specifying how they will be measured.
Our DV is participants' stated likelihood to make an appointment at their GP's office after reading a given scenario describing their health symptoms. Making an appointment involves either 1) exaggerating one's symptoms in an online form to secure the appointment (triage treatments) or 2) calling their GP's office to secure the appointment (non-triage treatments).
We only consider participants' choices in the 4 scenarios that have been vetted by our medical professional to be very unlikely to require a GP visit. We call these the low-harm scenarios. We note that each scenario has a low and high accuracy version where the high accuracy version provides a bit of additional information about the potential need for medical attention but is otherwise identical to the low accuracy version.

4) How many and which conditions will participants be assigned to?
We have 4 between subject treatments and 1 within subject treatment.

The 4 between subject treatments are as follows:
1) Baseline – participants are asked to state how likely they are to call their GP's office to secure an appointment after reading each scenario. There is no upfront cost of visiting one's GP's office and no need to exaggerate symptoms to secure an appointment.
2) Cost – identical to Baseline except that participants are aware that the (hypothetical) system is such that they would incur a £10 cost when visiting their GP.
3) Triage – identical to Baseline except instead of being asked how likely they are to call their GP's office to secure an appointment, they are asked how likely they are to exaggerate their symptoms in an online form to secure the appointment.
4) CostTriage – combines the Cost and Triage treatments such that participants not only incur the £10 cost when visiting their GP but also must exaggerate their symptoms to secure the appointment.

The within subject treatment is the Low-accuracy vs. High-accuracy versions of each scenario: Half the participants read the Low-accuracy version of 2 low-harm scenarios and the High-accuracy versions of the other 2 low-harm scenarios. Meanwhile, the other half of participants view the High-accuracy versions of the former and the Low-accuracy versions of the latter. There is only 1 high-harm scenario i.e., the scenario that does require medical attention. Half the participants view the High-accuracy version of it while the other half view the Low-accuracy version of it.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We have the following predictions for our DV, which is participants' stated likelihood to make an appointment at their GP's office. These are:

Prediction 1: Baseline > Cost
Prediction 2: Baseline > Triage
Prediction 3: Baseline > CostTriage
Prediction 4: Cost > CostTriage
Prediction 5: Triage > CostTriage

For Predictions 1-5, we conduct an OLS regression (Regression I), where we regress the DV on the following variables: (i) a dummy for each between-subject treatment (Cost, Triage, and CostTriage) and (ii) a dummy for the high and low accuracy version of each of the 4 low-harm scenarios, so in total 8 scenario dummies. We cluster standard errors at individual level. We check whether the estimates of each of the between-subject treatment dummies is significantly different from zero. In addition, we check whether the coefficients of these dummies are significantly different from each other using a Wald test.

Prediction 6: Base_Low > Base_High
Prediction 7: Cost_Low > Cost_High

Prediction 8: Triage_Low > Triage_High
Prediction 9: CostTriage_Low > CostTriage_High

For Predictions 6-9, we conduct an OLS regression (Regression II), where we regress the DV on the following variables: (i) a dummy for each between-subject treatment (Cost, Triage, and CostTriage), (ii) a dummy for each interaction of the between-subject treatment with the High-accuracy condition (Baseline_High, Cost_High, Triage_High, CostTriage_High), (iii) a dummy for each of 4 low-harm scenarios (i.e., regardless of whether it is the High-accuracy or Low-accuracy version). We cluster standard errors at individual level. We check whether the estimate of each of the dummies of (ii) is significantly different from zero.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will recruit a representative UK sample on Prolific. We will not invite participants who participated in a previous experiment that tested similar predictions using a different set-up. We exclude observations where participants did not complete the experiment or are not listed by Prolific as having completed the experiment.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We aim to collect data from 1200 participants in total.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
1. We check whether the estimates of the dummies (ii) in Regression II are significantly different from each other using a Wald test.

2. We will test for heterogeneous treatment effects based on age, gender, ethnicity, and income.

3. We will repeat the main analyses for the high-harm scenario. We acknowledge here that the interpretation of the outcome variable in the Triage treatment is less straightforward in this case. This is because for this treatment the outcome variable is participants' response to the question of how likely they are – to exaggerate their symptoms – to secure an appointment with their GP. Since our 14 medical professionals from the previous study judged this scenario as clearly warranting medical attention, strictly speaking, there should be no need to exaggerate one's symptoms to secure an appointment. We may therefore see a downward bias in responses in the Triage treatment depending on how participants interpret the question.

TABLE 10: Regression Results Vignette Experiment – First Displayed Scenario

	Dependent variable:	
	Reported likelihood of scheduling a GP appointment in %	
	Regression I	Regression II
	(1)	(2)
Cost	1.315	1.512
	(2.137)	(2.996)
Triage	4.873*	7.384*
	(2.209)	(3.116)
CostTriage	3.486	3.577
	(2.204)	(3.105)
Baseline × High accuracy		−5.876
		(3.024)
Cost × High accuracy		−6.292*
		(2.966)
Triage × High accuracy		−10.941***
		(3.199)
CostTriage × High accuracy		−6.106
		(3.178)
Constant	13.125***	20.288***
	(2.496)	(2.478)
Scenario Fixed	No	Yes
Scenario Accuracy Fixed	Yes	No
#Clusters (Individuals)	1200	1200
Observations	975	975
R^2	0.034	0.033
Adjusted R^2	0.024	0.023

*Note: SE clustered at individual level. * p<0.05, ** p<0.01, *** p<0.001*

TABLE 11: OLS regression: Likelihood to schedule a GP appointment

	Dependent variable:	
	Reported likelihood of scheduling a GP appointment in %	
	(1)	(2)
Cost	−0.64 (1.31)	−0.80 (1.54)
Triage	0.71 (1.45)	0.44 (1.60)
CostTriage	0.90 (1.35)	0.52 (1.54)
Female	−5.98*** (1.04)	−6.01*** (1.04)
Age: 31–40	−2.73 (1.65)	−2.70 (1.65)
Age: 41–50	−2.86 (1.79)	−2.80 (1.79)
Age: 51–60	−5.04*** (1.51)	−4.98** (1.51)
Age: 61+	−5.37*** (1.60)	−5.38*** (1.60)
Ethnicity: Asian	14.06*** (2.29)	14.11*** (2.29)
Ethnicity: Black	22.54*** (4.65)	22.60*** (4.66)
Ethnicity: Other	5.05 (3.15)	5.18 (3.15)
Income: £20,000 – £29,999	2.84* (1.43)	2.96* (1.42)
Income: £30,000 – £39,999	0.03 (1.41)	0.13 (1.40)
Income: £40,000 – £49,999	1.49 (1.97)	1.53 (1.97)
Income: £50,000 – £59,999	3.41 (2.13)	3.39 (2.14)
Income: £60,000 or more	0.32 (2.20)	0.39 (2.20)
Income: I'd rather not say	1.22 (2.38)	1.17 (2.38)
Baseline × High accuracy		−3.86*** (0.94)
Cost × High accuracy		−3.54*** (1.00)
Triage × High accuracy		−3.30*** (0.83)
CostTriage × High accuracy		−3.10*** (0.88)
Constant	23.33*** (1.98)	21.56*** (1.88)
Scenario Fixed	No	Yes
Scenario Accuracy Fixed	Yes	No
#Clusters (Individuals)	1,200	1,200
Observations	4,800	4,800
R^2	0.10	0.10
Adjusted R^2	0.10	0.09

*Note: SE clustered at individual level. * p<0.05, ** p<0.01, *** p<0.001*

TABLE 12: OLS regressions: Heterogeneous treatment effects by subgroup

	Dependent variable:							
	Reported likelihood of scheduling a GP appointment in %							
	Gender (1)	Gender (2)	Age (1)	Age (2)	Ethnicity (1)	Ethnicity (2)	Income (1)	Income (2)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cost	−0.22 (2.21)	−0.22 (1.65)	−0.65 (2.11)	−0.87 (1.66)	−5.46 (5.20)	−1.08 (1.60)	−0.51 (1.70)	−0.68 (1.67)
Triage	1.93 (2.52)	1.08 (1.69)	1.87 (2.31)	0.25 (1.68)	−7.48 (5.42)	0.05 (1.64)	−0.60 (1.76)	0.51 (1.69)
CostTriage	−0.60 (2.15)	0.66 (1.61)	0.02 (2.15)	0.23 (1.61)	−7.72 (5.21)	0.45 (1.56)	0.81 (1.68)	0.46 (1.61)
Subgroup	−6.96*** (2.01)	−6.28*** (1.20)	−4.71* (2.07)	−5.13*** (1.18)	−20.82*** (4.02)	−14.23*** (2.07)	1.52 (2.23)	3.14* (1.40)
Cost × Subgroup	0.38 (2.88)		−0.08 (2.92)		5.17 (5.37)		0.09 (3.31)	
Triage × Subgroup	−0.89 (3.16)		−2.70 (3.08)		9.09 (5.61)		5.41 (3.66)	
CostTriage × Subgroup	3.32 (2.84)		1.26 (2.89)		9.85 (5.39)		0.12 (3.32)	
Baseline × High accuracy		−3.80** (1.47)		−4.45** (1.46)		2.79 (3.91)		−2.79* (1.17)
Cost × High accuracy		−3.75* (1.57)		−3.67* (1.52)		−2.32 (3.46)		−3.37** (1.18)
Triage × High accuracy		−2.29 (1.57)		−1.84 (1.41)		−4.69 (3.08)		−4.55*** (1.08)
CostTriage × High accuracy		−3.78* (1.48)		−3.96** (1.37)		−5.33 (3.34)		−2.57* (1.13)
Baseline × High accuracy × Subgroup		−0.12 (2.17)		1.13 (2.20)		−7.65 (4.24)		−4.08 (2.40)
Cost × High accuracy × Subgroup		0.39 (2.11)		0.27 (2.07)		−1.45 (3.73)		−0.61 (2.44)
Triage × High accuracy × Subgroup		−1.80 (2.27)		−3.02 (2.17)		1.65 (3.57)		4.90 (2.70)
CostTriage × High accuracy × Subgroup		1.35 (2.05)		1.75 (2.06)		2.57 (3.69)		−2.07 (2.33)
Constant	23.41*** (1.70)	21.21*** (1.35)	22.69*** (1.79)	20.98*** (1.40)	38.22*** (4.00)	30.61*** (2.25)	19.82*** (1.51)	17.46*** (1.29)
Scenario Fixed	No	Yes	No	Yes	No	Yes	No	Yes
Scenario Accuracy Fixed	Yes	No	Yes	No	Yes	No	Yes	No
#Clusters (Individuals)	1,200	1,200	1,200	1,200	1,200	1,200	1,200	1,200
Observations	4,800	4,800	4,800	4,800	4,800	4,800	4,800	4,800
R^2	0.03	0.03	0.03	0.02	0.07	0.06	0.02	0.01
Adjusted R^2	0.03	0.02	0.02	0.02	0.06	0.06	0.01	0.01

Note: The Subgroup dummy is a dummy for either Gender: Female in (1) & (2), Age: Above 48 Years in (3) & (4), Ethnicity: White in (5) & (6), and Income: Above £40000 in (7) & (8). Robust SEs clustered by participant

*p<0.05; **p<0.01; ***p<0.001

C.D Validation of Health Scenarios in Experiment 2

We first developed eight health-related scenarios in consultation with medical professionals. These scenarios were subsequently validated in a Prolific study with 14 certified UK general practitioners (GPs). Based on this validation, we selected three scenarios out of these eight: (ii) Scenario 1, with an average rating of 4.5 on a 0–10 scale (where 0 = "definitely do not visit a GP"), labeled as Scenario A; (ii) Scenario 4, with an average rating of 4.0 (Scenario B); and (iii) Scenario 6 (low accuracy), which received an average GP recommendation of 2.7 (Scenario C).[25] For Scenarios 1 and 4, we merged the high- and low-accuracy versions. For all three selected scenarios, we then constructed new high-accuracy variants. Importantly, all selected scenarios had mean ratings below 5.0, indicating that GPs, on average, advised against a visit. At the same time, there was substantial heterogeneity in GP evaluations, which we consider advantageous, as the scenarios were intended to provide noisy but informative signals consistent with low harm. In addition, we developed one further scenario D, which 8 out of 14 GPs identified as a common situation in which patients unnecessarily visit their GP. Finally, we included a high-harm scenario 9 in which a GP visit is unequivocally warranted (Scenario E). This scenario received an average recommendation of 9.57 out of 10.

Another aim of the validation study was to obtain empirical estimates of how frequently individuals visit their GP under these types of scenarios. All 14 GPs agreed—10 strongly (7 on a 7-point Likert scale) and 4 somewhat (6 on a 7-point Likert scale)—that "sometimes, individuals may make appointments to see their GP despite their symptoms not really being ones that require medical attention."

C.E Analysis of Participants' Open-text Responses

In order to evaluate participants' reasoning for scheduling (or not scheduling) a GP appointment, we make use of their open-text responses to the post-treatment questionnaires, in which they explain the factors they considered when making their decisions. We analyze these textual data using computational text analysis techniques in two steps: first, we identify the most prevalent words in participants' responses; second, we apply dictionary-based analysis to examine whether responses contain terms that are expected to be associated with particular motivations.

First, we examine the most frequently used words within each treatment using term-frequency analysis. To ensure that the resulting terms are meaningful for our purposes, we pre-process participants' responses by lower-casing the text, removing punctuation and numbers, tokenizing responses into individual words, lemmatizing words to their dictionary form[26], removing stop words, and excluding words shorter than three characters. We then compute word frequencies separately for each treatment and visualize the 15 most frequent terms in Figure 8. It shows that across all treatments, the word *symptom* is by far the most frequently used term. Terms related to cost or exaggerate do appear in the respective treatments, yet much less frequently than symptoms.

[25] The averages are calculated across both low- and high-accuracy versions of these scenarios.

[26] For instance, after lemmatization, both *exaggerated* and *exaggerating* are reduced to *exaggerate*.

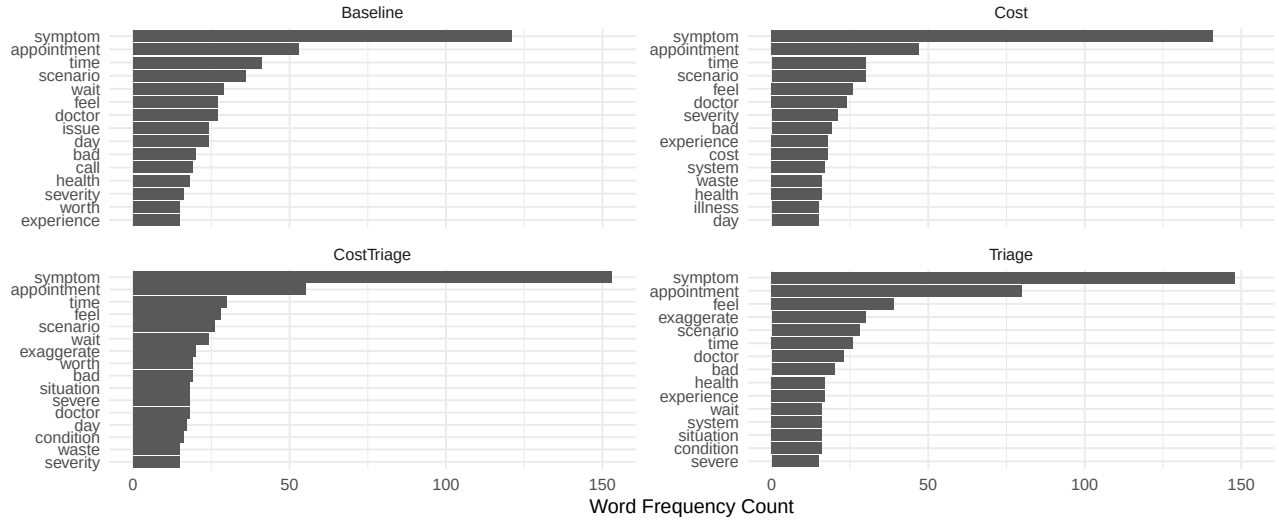


FIGURE 8: Frequency of Words in Reasons for Extraction per Treatment

Second, we apply a dictionary-based classification approach to quantify the prevalence of different types of considerations mentioned in participants' responses. We construct five dictionaries capturing (i) symptom-related severity and urgency, (ii) health care system-related concerns, (iii) non-monetary payoff considerations, (iv) references to exaggeration or honesty, and (v) monetary cost considerations.[27] For this analysis, pre-processing is kept to a minimum and includes only lower-casing and the removal of punctuation. Each response is then evaluated for matches with the category dictionaries, and we record a binary indicator for whether at least one term from a given category occurs in the response. For each category, we calculate the proportion of participants whose responses contain at least one such term and summarize these measures by treatment using 95% Wilson binomial confidence intervals in Figure 9. The figure illustrates that across all treatments, approximately 70% of participants justify their decision to schedule (or not schedule) a GP visit by referring to symptom-related considerations (e.g., symptom, severity, urgency, ill). Only about 10% additionally cite system-related concerns, e.g., the overwhelmed system, not wanting to burden the GP or the healthcare system. In the *Cost* treatments, roughly 10% mention the monetary cost, often noting, however, that it did not affect their decision. Similarly, in the *Triage* treatments, only around 10% refer to the need to exaggerate symptoms as part (or not part) of their decision-making.

[27] The symptom-related severity dictionary consists of the following terms: *symptom, severity, severe, serious, serios, ill, consequence, urgent, persist, i needed, i really needed, i truly needed, long, condition, life threatening, experience, go away, worried, risk, importance, sick, concerning, necessary, information, unwell, damage, pain, cause, bad.* The system-related dictionary includes: *overwhelmed, burden, GP's time, doctor's time, physician's time, their time, anyone's time, waste, someone else, others, trouble the, troubling the, appropriate, health service, warrant an appointment, warranted an appointment, more urgent, embarrassment, deserv, busy, drain, justif, overworked, merit, healthcare system.* The non-monetary payoff dictionary includes: *hassle, waiting time, worth it, waiting room, treatable, effort, my time, anyone's time.* The exaggeration and honesty dictionary includes: *exaggerat, fair, honest.* The monetary cost dictionary includes: *money, cost, charge, not cost, pounds 10, fee, price.*

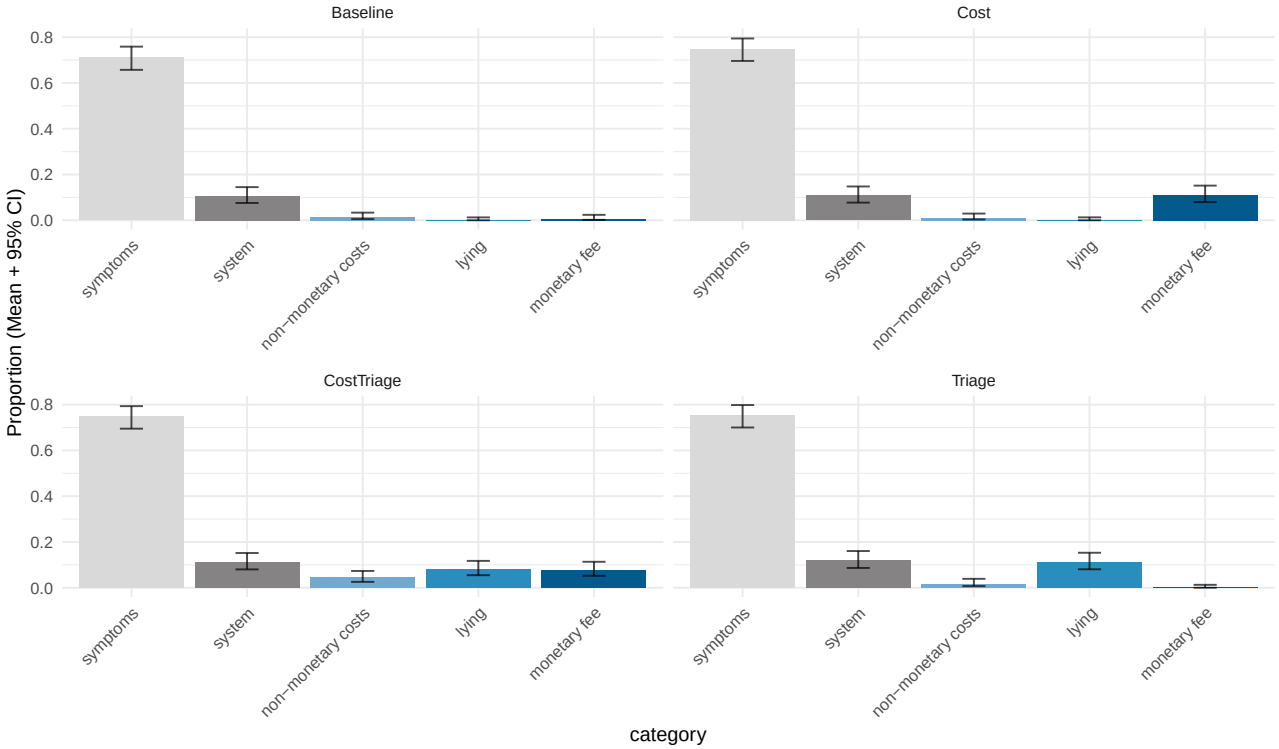
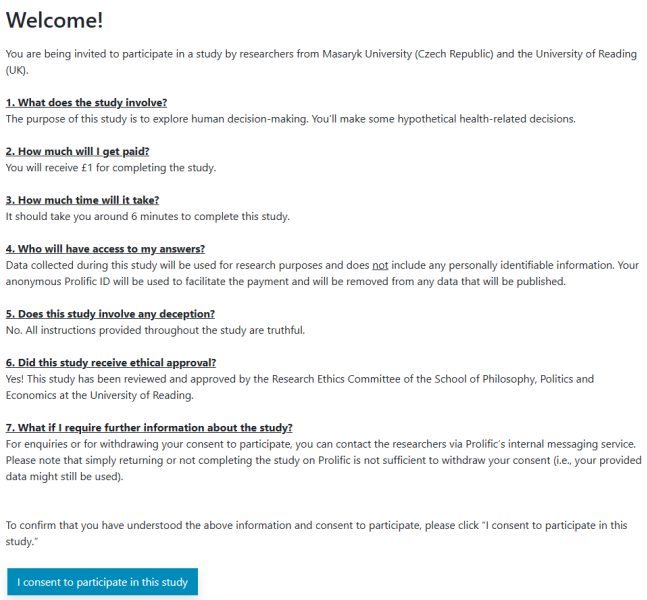


FIGURE 9: Frequency of Words in Reasons for Extraction per Treatment

C.F Instructions

This section presents screenshots of the environment used in Experiment 2, in the sequence in which they were displayed to participants. The screens correspond to the *Baseline* treatment and scenario combination T1, in which Scenarios A and C are presented in their high-accuracy versions, while Scenarios B, D, and E are shown in their low-accuracy versions. The full texts of all scenarios are reported in Table 13. Instructions in the other strictly between-subject treatments differ primarily in Screen 2 (Welcome) and Screens 4–8 (Scenarios); we therefore present Screen 2 and Screen 4 in all treatment-specific variants. Screens 5–8 are shown only for the Baseline treatment, but would vary analogously across treatments.



Welcome!

You are being invited to participate in a study by researchers from Masaryk University (Czech Republic) and the University of Reading (UK).

1. What does the study involve?
The purpose of this study is to explore human decision-making. You'll make some hypothetical health-related decisions.

2. How much will I get paid?
You will receive £1 for completing the study.

3. How much time will it take?
It should take you around 6 minutes to complete this study.

4. Who will have access to my answers?
Data collected during this study will be used for research purposes and does not include any personally identifiable information. Your anonymous Prolific ID will be used to facilitate the payment and will be removed from any data that will be published.

5. Does this study involve any deception?
No. All instructions provided throughout the study are truthful.

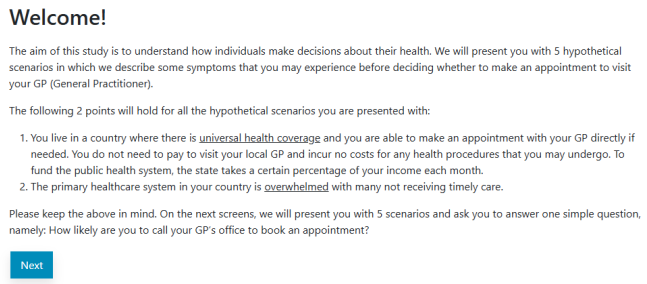
6. Did this study receive ethical approval?
Yes! This study has been reviewed and approved by the Research Ethics Committee of the School of Philosophy, Politics and Economics at the University of Reading.

7. What if I require further information about the study?
For enquiries or for withdrawing your consent to participate, you can contact the researchers via Prolific's internal messaging service. Please note that simply returning or not completing the study on Prolific is not sufficient to withdraw your consent (i.e., your provided data might still be used).

To confirm that you have understood the above information and consent to participate, please click "I consent to participate in this study."

I consent to participate in this study

Screen 1



Welcome!

The aim of this study is to understand how individuals make decisions about their health. We will present you with 5 hypothetical scenarios in which we describe some symptoms that you may experience before deciding whether to make an appointment to visit your GP (General Practitioner).

The following 2 points will hold for all the hypothetical scenarios you are presented with:

1. You live in a country where there is universal health coverage and you are able to make an appointment with your GP directly if needed. You do not need to pay to visit your local GP and incur no costs for any health procedures that you may undergo. To fund the public health system, the state takes a certain percentage of your income each month.
2. The primary healthcare system in your country is overwhelmed with many not receiving timely care.

Please keep the above in mind. On the next screens, we will present you with 5 scenarios and ask you to answer one simple question, namely: How likely are you to call your GP's office to book an appointment?

Next

Screen 2 – Baseline

Welcome!

The aim of this study is to understand how individuals make decisions about their health. We will present you with 5 hypothetical scenarios in which we describe some symptoms that you may experience before deciding whether to make an appointment to visit your GP (General Practitioner).

The following 2 points will hold for all the hypothetical scenarios you are presented with:

1. You live in a country where there is universal health coverage and you are able to make an appointment with your GP directly if needed. You still incur a nominal cost of £10 each time you visit your local GP but incur no costs for any health procedures that you may undergo. To fund the public health system, the state takes a certain percentage of your income each month.
2. The primary healthcare system in your country is overwhelmed with many not receiving timely care.

Please keep the above in mind. On the next screens, we will present you with 5 scenarios and ask you to answer one simple question, namely: How likely are you to call your GP's office to book an appointment?

Next

Screen 2 – Cost

Welcome!

The aim of this study is to understand how individuals make decisions about their health. We will present you with 5 hypothetical scenarios in which we describe some symptoms that you may experience before deciding whether to make an appointment to visit your GP (General Practitioner).

The following 3 points will hold for all the hypothetical scenarios you are presented with:

1. You live in a country where there is universal health coverage and you are able to make an appointment with your GP directly if needed. You do not need to pay to visit your local GP and incur no costs for any health procedures that you may undergo. To fund the public health system, the state takes a certain percentage of your income each month.
2. The primary healthcare system in your country is overwhelmed with many not receiving timely care.
3. To get an appointment at your local GP, you must first complete an online form describing your symptoms and only patients who report severe symptoms will get an appointment. That said, you are free to exaggerate your symptoms to guarantee yourself an appointment.

Please keep the above in mind. On the next screens, we will present you with 5 scenarios and ask you to answer one simple question, namely: How likely are you to exaggerate your symptoms in your GP's online form so as to book an appointment?

Next

Screen 2 – Triage

Welcome!

The aim of this study is to understand how individuals make decisions about their health. We will present you with 5 hypothetical scenarios in which we describe some symptoms that you may experience before deciding whether to make an appointment to visit your GP (General Practitioner).

The following 3 points will hold for all the hypothetical scenarios you are presented with:

1. You live in a country where there is universal health coverage and you are able to make an appointment with your GP directly if needed. You still incur a nominal cost of £10 each time you visit your local GP but incur no costs for any health procedures that you may undergo. To fund the public health system, the state takes a certain percentage of your income each month.
2. The primary healthcare system in your country is overwhelmed with many not receiving timely care.
3. To get an appointment at your local GP, you must first complete an online form describing your symptoms and only patients who report severe symptoms will get an appointment. That said, you are free to exaggerate your symptoms to guarantee yourself an appointment.

Please keep the above in mind. On the next screens, we will present you with 5 scenarios and ask you to answer one simple question, namely: How likely are you to exaggerate your symptoms in your GP's online form so as to book an appointment?

Next

Screen 2 – CostTriage

Comprehension Questions

Before you begin, please answer all the questions about the described scenario.

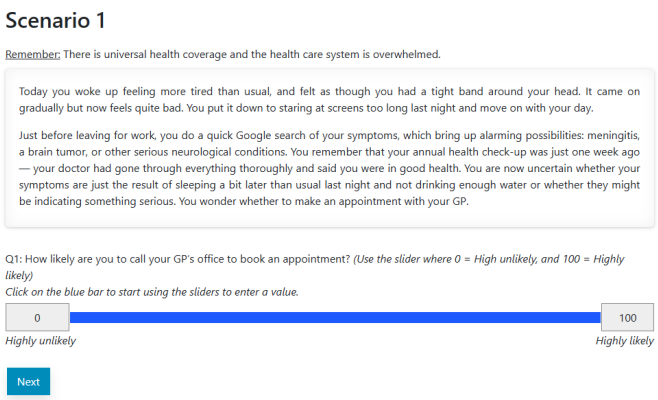
Q1: How much do you need to pay when you visit your local GP?

Q2: What is the status of the healthcare system in your country?

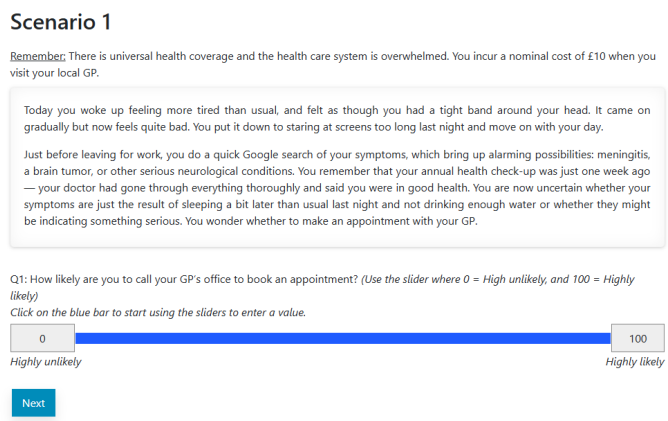
Q3: How many scenarios will you be presented with on the next screens?

Next

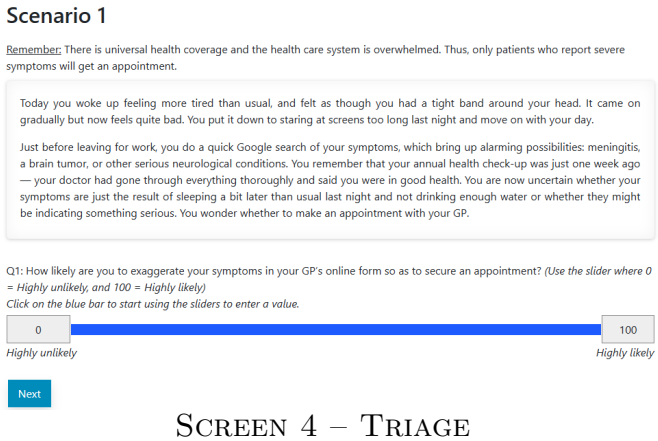
Screen 3



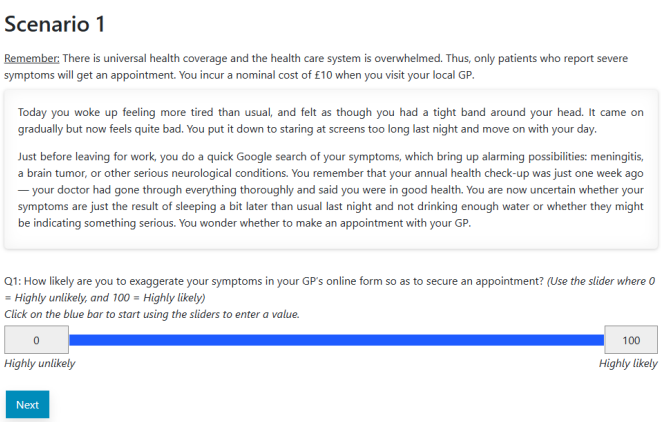
Screen 4 – Baseline



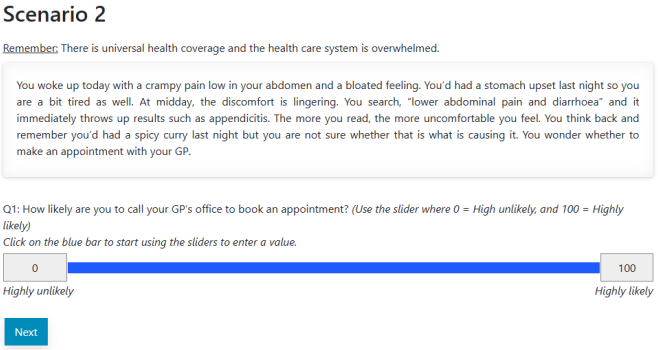
Screen 4 – Cost



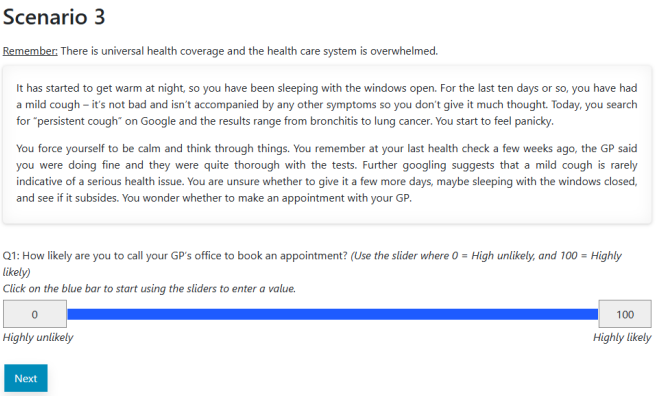
Screen 4 – Triage



Screen 4 – CostTriage



Screen 5



Screen 6

Scenario 4

Remember: There is universal health coverage and the health care system is overwhelmed.

You have just returned from a holiday, after visiting your brother and your 2-year-old niece. Your niece was down with a viral infection, so you didn't have much fun. Today you wake up feeling quite unwell with a sore throat and a cough. You don't bring up any phlegm but the cough persists the whole day. You stay in bed most of the day but by the end of the day you feel extremely tired and your muscles ache. You wonder whether to make an appointment with your GP.

Q1: How likely are you to call your GP's office to book an appointment? *(Use the slider where 0 = High unlikely, and 100 = Highly likely)*
Click on the blue bar to start using the sliders to enter a value.

0 — 100

Highly unlikely — *Highly likely*

Next

SCREEN 7

Scenario 5

Remember: There is universal health coverage and the health care system is overwhelmed.

You have been feeling a bit off for the past week — some nausea here and there, a vague sense of discomfort in your upper abdomen after eating, and more recently, a dull ache in your upper back that comes and goes. This morning, after climbing a single flight of stairs, you feel unexpectedly short of breath. You are not wheezing, but you are winded in a way that doesn't feel normal. Later, when sitting down at your desk, the discomfort comes back with a wave of nausea and cold sweat.

You google your symptoms. To your surprise, several medical websites warn that these could be signs of a "silent" heart attack even if the pain isn't sharp or dramatic. That sounds extreme to you — nothing feels urgent exactly — but now you are sitting still and your heart seems to be beating harder than it should.

You wonder whether it may make sense to make an appointment with your GP.

Q1: How likely are you to call your GP's office to book an appointment? *(Use the slider where 0 = High unlikely, and 100 = Highly likely)*
Click on the blue bar to start using the sliders to enter a value.

0 — 100

Highly unlikely — *Highly likely*

Next

SCREEN 8

Questionnaire

Please answer some final questions below and click "Submit" to complete the study.

Q1. What is your gender?

Q2. What is your age (in years)?

Q3. What is your approximate annual income (before tax)?

Q4. What is your ethnicity?

Q5. What were you thinking about when you were deciding whether to make an appointment with your GP?

Q6. Do you have any comments?

Submit

SCREEN 9

C.G Accuracy Versions of Scenarios

TABLE 13: High-accuracy versions of scenarios used in Experiment 2. Italics indicate elements specific to the high-accuracy versions.

Scenario A	Today you woke up feeling more tired than usual, and felt as though you had a tight band around your head. It came on gradually but now feels quite bad. You put it down to staring at screens too long last night and move on with your day. Just before leaving for work, you do a quick Google search of your symptoms, which bring up alarming possibilities: meningitis, a brain tumor, or other serious neurological conditions. *You remember that your annual health check was just one week ago—your doctor had gone through everything thoroughly and said you were in good health.* You are now uncertain whether your symptoms are just the result of sleeping a bit later than usual last night and not drinking enough water or whether they might be indicating something serious. You wonder whether to make an appointment with your GP.
Scenario B	You woke up today with a crampy pain low in your abdomen and a bloated feeling. You'd had a stomach upset last night so you are a bit tired as well. At midday, the discomfort is lingering. You search, "lower abdominal pain and diarrhoea" and it immediately shows up results such as appendicitis. The more you read, the more uncomfortable you feel. You think back and remember you had a spicy curry last night but you are not sure whether that is what is causing it. *You also recall that you barely drank any water yesterday and that you have actually experienced similar symptoms once before and it passed on its own.* You wonder whether to make an appointment with your GP.
Scenario C	It has started to get warm at night, so you have been sleeping with the windows open. For the last ten days or so, you have had a mild cough—it's not bad and isn't accompanied by any other symptoms so you don't give it much thought. Today, you search for "persistent cough" on Google and the results range from bronchitis to lung cancer. You start to feel panicky. *You force yourself to be calm and think through things. You remember at your last health check a few weeks ago, the GP said you were doing fine and they were quite thorough with the tests. Further googling suggests that a mild cough is rarely indicative of a serious health issue.* You are unsure whether to give it a few more days, maybe sleeping with the windows closed, and see if it subsides. You wonder whether to make an appointment with your GP.
Scenario D	You have just returned from a holiday, after visiting your brother and your 2-year-old niece. Your niece was down with a viral infection, so you didn't have much fun. Today you wake up feeling quite unwell with a sore throat and a cough. You don't bring up any phlegm but the cough persists the whole day. You stay in bed most of the day but by the end of the day you feel extremely tired and your muscles ache. *You lay in bed and try to think about the last few days. You realize that you most likely caught the viral infection your niece had and in that case, you know that nothing much can be done except to wait it out. Still, you don't feel well at all and* [Y]ou wonder whether you should make an appointment to see your GP.
Scenario E	You have been feeling a bit off for the past week—some nausea here and there, a vague sense of discomfort in your upper abdomen after eating, and more recently, a dull ache in your upper back that comes and goes. This morning, after climbing a single flight of stairs, you feel unexpectedly short of breath. You are not wheezing, but you are winded in a way that doesn't feel normal. Later, when sitting down at your desk, the discomfort comes back with a wave of nausea and cold sweat. You google your symptoms. To your surprise, several medical websites warn that these could be signs of a "silent" heart attack even if the pain isn't sharp or dramatic. That sounds extreme to you—nothing feels urgent exactly—but now you are sitting still and your heart seems to be beating harder than it should. *You call a friend in the medical profession and describe how you have been feeling. They ask you a few questions and you notice that their tone has changed—it has become more serious. They say that you must likely need to see a doctor and sooner rather than later. You know your friend to be extra cautious so when you put down the phone,* [Y]ou wonder *again* whether it may make sense to make an appointment with your GP.