Stated preferences outperform elicited preferences for predicting reported compliance with Covid-19 prophylactic measures

Ismaël Rafaï, Thierry Blayac, Dimitri Dubois, Sébastien Duchêne, Phu Nguyen-Van, Bruno Ventelou, Marc Willinger

Stated preferences outperform elicited preferences for predicting reported compliance with Covid-19 prophylactic measures.

Ismaël Rafaï (1,2,*), Thierry Blayac (1), Dimitri Dubois (1), Sébastien Duchêne (3), Phu Nguyen-Van (4,5,6), Bruno Ventelou (2), Marc Willinger (1)

Abstract

This paper studies the behavioral and socio-demographic determinants of reported compliance with prophylactic measures against COVID-19: barrier gestures, lockdown restrictions and mask wearing. The study contrasts two types of measures for behavioral determinants: experimentally elicited preferences (risk tolerance, time preferences, social value orientation and cooperativeness) and stated preferences (risk tolerance, time preferences, and the GSS trust question). Data were collected from a representative sample of the metropolitan French adult population (N=1154) surveyed during the first lockdown in May 2020, and the experimental tasks were carried out on-line. The in-sample and out-of-sample predictive power of several regression models - which vary in the set of variables that they include - are studied and compared. Overall, we find that stated preferences are better predictors of compliance with these prophylactic measures than preferences elicited through incentivized experiments: self-reported level of risk, patience and trust are predicting compliance, while elicited measures of risk-aversion, patience, cooperation and prosociality did not.

Keywords: COVID-19, individual preferences, social preferences, elicited preferences, stated preferences.

JEL: C90, D90, I18

Acknowledgments: we acknowledge financial support from the national program AAP "Flash COVID-19" from ANR (ANR-20-COVI-000) and from the regional program "Défis clés Urgence COVID-19" from the "Région Occitanie".

(*) corresponding author : ismael.rafai@gmail.com
(1) CEE-M, Univ Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France
(2) Aix Marseille Univ, CNRS, AMSE, Marseille, France
(3) Montpellier Business School, Montpellier, France
(4) ECONOMIX, CNRS, University of Paris Nanterre, Nanterre (France);
(5) University of Management and Technology, Ho Chi Minh City (Vietnam);
(6) FERDI, Clermont-Ferrand (France)

1. Introduction

The COVID-19 pandemic constitutes an unprecedented global event. To resolve the crisis, government responses are based on two complementary strategies: 1) limiting the danger posed by the disease by supporting the medical sector and developing appropriate treatments, 2) reducing the spread of the disease by means of prophylactic measures. This paper focuses on the second strategy, which rests on individual willingness to follow the authorities' recommendations and restrictions in terms of barrier gestures (e.g., washing hands, avoiding touching face, coughing in sleeves, and social distancing), mask wearing, and lockdown compliance. Individual decisions to comply with these measures can be modeled as a trade-off involving temporal, risk and social dimensions. Indeed, respecting these prophylactic measures is costly for the individuals, who must change their daily habits by washing their hands more often, avoiding touching their face, and coughing in their sleeves; must limit their social interactions by social distancing and respecting lockdown; and must suffer discomfort by wearing a mask (Labiris et al. 2021). However, compliance reduces self-risk of a future infection. Due to the contagious nature of the disease, compliance induces positive externalities and contributes to the mitigation of the average collective rate of infection.

Understanding the source of heterogeneity in rates of compliance with prophylactic measures is essential to the proposal of adequate policies (Geoffard and Philipson, 1996). From an economic point of view, compliance heterogeneity can be explained by two sets of factors: individuals' characteristics (e.g. gender or age) and individuals' preferences (e.g. risk, time and social preferences). Firstly, individuals' characteristics shape their compliance costs and utility: for instance, lockdown compliance may be costlier for people living in smaller apartments, and being infected may have more serious consequences for persons suffering from comorbidities such as diabetes or high blood pressure. Secondly, individuals' personality traits and their risk, time and social preferences shape the way in which the immediate personal costs incurred by compliance with some prophylactic measures are discounted by the individual, compared to the uncertain future costs of being infected or infecting a relative.

The first set of factors, which includes age, gender, chronic conditions, education, political orientation, welfare, civil culture, and trust in science and medicine, has been studied extensively in the fast growing body of literature on Covid-19 (see e.g. Barrios et al., 2021; Bertin et al., 2020; Durante et al., 2021; Gadarian et al., 2021; Galasso et al., 2020; Nivette et al., 2020; Painter and Qiu, 2021; Plohl and Musil., 2021; Sailer et al., 2022; Szabo et al., 2020; Wright et al., 2020), as has the importance of personality traits such as empathy, impulsivity, amorality, egoism, or psychopathy (Kuiper et al., 2020; Zajenowski et al., 2020; Zettler et al., 2022). In this paper we focus on the behavioral determinants of compliance; these consist of both self-centered preferences (e.g. risk and time preferences) and social preferences (e.g. trust, cooperativeness and prosociality).

Consider how self-centered preferences might affect compliance. Compliance is costly for individuals, as it reduces their freedom of movement (lockdown), implies discomfort (mask wearing), or changes their habits (barrier gestures), but it is beneficial for the individual and for the collective, as it reduces viral exposure and thereby the likelihood of infection. Therefore, it is likely that more risk-averse individuals are more willing to comply in order to reduce their exposure to adverse events. By contrast, individuals who are more impatient are likely to be less compliant, because the immediate cost looms larger in their minds than the discounted benefits. In terms of other-regarding preferences, more cooperative individuals are expected to be more likely to comply with constraining measures, as these more prosocial individuals should be more inclined to avoid infecting others (Galizzi and Navarro-Martinez, 2019).

This rationale could be illustrated by considering a binary decision where the individual chooses to comply or not with the recommendations to reduce the future infection probability. Compliance is expected to be profitable for the individual if and only if the ratio between compliance costs (i.e. the direct loss in utility when following the recommendations) and discounted infection costs (i.e. the discounted loss of utility from being infected) is lower than the reduction in the infection probability due to compliance (Δp). It follows that more risk adverse individuals should be more likely to comply. Indeed (since being infected is worse than complying) an increase in utility function concavity decreases compliance costs more that it decreases infection costs, thus decreasing the ratio between the two. Similarly, more patient and more prosocial individuals should be more likely to comply since prosocial preferences increases infection costs and patience increases the discount factor. Meanwhile, truthfulness may have an ambiguous effect on compliance. People who trust more in others might hold false beliefs about others' trustworthiness. For example, they might believe that their peers strongly respect barrier gestures (i.e. Δp is considered as lower), which could lead them to adopt a lax attitude towards lockdown. On the other hand, trust might be targeted towards believing that others adopt a compliant attitude, which would lead the trustful to imitate the expected norm of compliance (e.g. if we consider belief-dependent utility, higher trust reduces the compliance costs).

In contrast to individuals' characteristics, preferences are unobservable; this raises an important issue about their appropriate measurement. The two standard approaches are the stated preferences method (SP thereafter), which is based on questionnaires, and the experimentally elicited preferences method (EP thereafter), which is based on incentivized tasks. Many researchers claim that EP methods are preferable because they rely on tasks that are designed to reveal respondents' preferences in an incentive compatible way. By relying on the (weak) axiom of revealed preferences of consumer theory

(Samuelson, 1938), EP methods supposedly allow the researcher to track true preferences more effectively. By contrast, the SP method is not based on any axiomatic or theoretical framework. The strongest criticism of the SP method is with regard to its hypothetical bias, which is a key challenge for the provision of credible willingness-to-pay measurements in contingent valuation studies (List and Gallet 2001; Carlsson, 2010). A considerable amount of effort has been devoted to circumventing this criticism by designing "smart truth-serums", such as honesty priming (de-Magistris et al., 2013; Howard et al., 2017), preference elicitation under oath (Jacquemet et al., 2013) or under the ten commandments (Mazar et al., 2008; Lim et al., 2015), or other sophisticated tools (see e.g. de Corte et al., 2021). Besides, there seems to be an intrinsic preference for truth-telling (Abeler et al., 2019), which may dispense from using sophisticated tricks to get closer to the truth. Nevertheless, the relevance of the SP method remains an open question. Unlike the parsimonious and context-free laboratory tasks involved in EP, SP methods are exposed to a multiplicity of possible combinations of questions, framings and contexts, without any theoretical framework to guide the researcher.

Despite these limitations, several papers have shown that SP methods tend to outperform EP methods, particularly with regard to risk preferences; this casts serious doubts on the external validity of EP methods, i.e. their ability to make predictions which are relevant beyond lab behavior (Frey et al., 2017; Hertwig et al., 2019; Charness et al., 2020; Hertwig et al., 2019). The stability of various preference measures has also been questioned, with regard to both EP tasks (Pedroni et al., 2017; Holzmeister and Stefan, 2021) and methods (e.g., Craig et al., 2017; Fossen and Clocker, 2017; Chuang and Schechter, 2015). Today, the superiority of EP methods in getting closer to the true preferences of individuals is hotly debated (Arslan et al., 2020). A recent stream of empirical literature therefore relies heavily on SP methods (Dohmen et al., 2011; Falk et al., 2016; Falk et al., 2018). In this article, we thus address the question of whether EP or SP can predict reported compliance with COVID-19 prophylactic measures. We do this for both individual preferences (risk and time preferences) and social preferences (trust, cooperativeness, and prosociality).

The role of risk, time and social preferences in compliance with COVID-19 prophylactic measures has been empirically investigated in several recent papers. Using a non-incentivized survey on a German student population (N=185), Müller and Rau (2021) found that more patient and more risk-averse students expressed higher compliance in the COVID-19 context. A study based on a sample of US undergraduates (N=338) by Sheth and Wright (2020) did not find an effect of self-reported level of willingness to take risk and willingness to give, as measured using the Global Preference Survey (GPS hereafter, Falk et al., 2018), on the self-reported level of compliance with the recommendation in California to stay at home. The paper most similar to ours is that of Campos-Mercade *et al.* (2021), in which 13 different behaviors likely to reduce the spread of the COVID-19 were studied among a representative (in terms of age, gender, and county residency) Swedish sample (N=967), with various self-centered and other-regarding preferences measured using incentivized tasks and GPS questions. However, the scope of their paper was to test whether more prosocial individuals were more compliant. In this paper, we focus on the predictive power of stated and elicited preferences measurement methods with respect to (reported) compliance with barrier gestures, lockdown and mask wearing, and contrast the results between the type of methods. Another innovation in our paper is the use of several statistical methods which allow us to investigate both the in-sample predictive power (i.e. how precisely the model fits the data of the sample it is estimated from) and the out-sample predictive power (i.e. how precisely the model fits the data from other samples).

Our paper contributes both to the literature on the external validity of economic preferences measures, in this case on the respective abilities of SP and EP measures to predict reported field behavior, and to the behavioral literature on COVID-19. The key features of our research are as follows: (i) the study is based on a large sample of the metropolitan French adult population which is representative in terms of age, gender, region, and household income; (ii) it embraces an extended set of preferences; (iii) it address the respective predictive power of standard EP and SP measures.

More precisely, we focused on a set of prophylactic measures which included "washing hands", "avoiding touching the face", "social distancing", "coughing into sleeves", "mask wearing", and "respecting the lockdown". For each of these measures, we collected a level of compliance self-reported by individuals and respectively assess the predictive power of SP (based on validated questionnaires) and that of EP (based on incentivized tasks). The incentivized tasks allowed the elicitation of risk preferences using Gneezy and Potters' (1997) portfolio choice task, time preferences using the Convex Time Budget method (Andreoni and Sprenger, 2012), prosociality based on the Social Value Orientation (SVO) task (Murphy et al., 2011), and cooperativeness measured by the level of voluntary contribution to a linear public good. With regard to SP, standard questionnaires were used to measure risk preferences, self-reported patience (Dohmen et al., 2011, Falk et al., 2016), and trust (General Social Survey).

In order to assess the predictive power of each preference measurement for compliance, we relied on two statistical strategies. The first strategy consisted of applying a nested model comparison scheme to determine, across the whole sample, whether adding one type of preference measurement (EP and SP) significantly increased the goodness of fit of the models describing reported compliance with prophylactic measures. The second strategy was based on the method proposed by Ellies-Oury et al. (2019). This method, using a train/test sample procedure inspired by machine learning, aims to choose, among several regression models and with a selection of a limited number of explanatory variables, the one predicting reported compliance the best. This last procedure allowed us to assess out-of-sample predictive power (the ability of a model to predict observations that have not been used in the model estimation) and is less prone to over-fitting. Different regression models were tested, including linear regression, ordered logit regression, sliced inverse regression, random forest, principal component regression, partial least square regression, and ridge regression.

Our main finding – which is robusts to the statistical strategy -- is that our EP measures, including i.e. risk aversion, impatience, cooperativeness and prosociality, are poor predictors of declared compliance with barrier gestures, lockdown and mask wearing. Only one of these variables (cooperativeness) was correlated at a conventional significance level with only one of the recommended prophylactic measures ("avoiding touching the face"). Moreover, taken together, the EP set did not significantly increase the goodness of fit of our models and adding these variables did not increase the out-of-sample predictive power of most of the models. By contrast, our measures of SP were predictors of self-declared compliance: for instance, self-reported willingness to take risk (in general and in the health domain) were negatively correlated with compliance with prophylactic measures, and higher stated patience positively affected lockdown compliance. Stated trust, however, had ambiguous effects that are not easy to interpret.

Besides our main result concerning behavioral determinants, we report additional findings about socio-demographic determinants, most of which agree with previous findings in the literature. Men were less compliant than women, young respondents were less compliant with respect to social distancing, and elders reported washing their hands and coughing in their sleeves less often than others. However, elders also reported higher compliance with respect to social distancing and mask wearing. The virus prevalence rate in the respondents' region was positively correlated with the reported level of lockdown compliance.

The contrast between the predictive power of SP and EP measures in predicting reported compliance in our sample question the external validity of EP measures. A possible reason is that SP methods provide better predictors for stated (self-reported) behaviors, and that, symmetrically, EP methods would be better predictors for real-life health-related behaviors than SP methods. Indeed, an important limitation of our study is the fact that we were not able to directly observe compliance and relied instead on stated compliance. It is therefore likely that SP are better predictors, simply because both stated behavior and stated preferences pertain to a common cognitive process that converts actual preferences and behaviors into statements made by the same respondents, in particular during the same survey.

We address this issue in the discussion section.

In section 2 we describe our methodology. Section 3 presents the statistical results, which are then discussed in section 4.

2. Data and Methods

2.1. Design of the survey: sample and representativeness

We conducted an online experiment on a representative sample (N=1154) of the metropolitan French population between May 4th and May 16th, 2020. Respondents were recruited by phone at the end of March 2020 by the survey institute Viavoice[1]. Of the 7500 persons contacted by phone, 5331 accepted the invitation to participate and received a web link; 1154 of these fully completed the survey and signed the informed consent form (a response rate of 21.6%). The survey was implemented using the oTree platform (Chen et al. 2016), on a dedicated server managed by the research team. The sample was representative of the metropolitan French population in terms of gender, age distribution, and living area (see Appendix). This survey was part of a larger project, with variables also collected for other purposes and presented in other papers.[2]

We first asked participants to report their level of compliance with prophylactic measures (section 2.2), second, they participated to incentivized games (section 2.3), then participants completed a survey including personal information (see controls in section 2.4) followed by measures of stated preferences (section 2.3).

2.2. Prophylactic measures

In the survey, we asked respondents to report their actual compliance with prophylactic measures including barrier gestures (washing hands, coughing into sleeves, respecting social distance, avoiding touching one's face), lockdown compliance, and mask wearing.

Compliance to barrier gestures was measured by the following questions:

"During the lockdown, did you respect the following recommendations when you left your home?"

- *Washing your hands. (1 = "Never", 2 = "Sometimes", 3 = "Often" ,4 = "Very Often", 0 = "I don't know")*

- *Coughing in your sleeves. (1 = "Never", 2 = "Sometimes", 3 = "Often" ,4 = "Very Often", 0= "I don't know")*

- *Respecting a distance of at least one meter with other people. (1 = "Never", 2 = "Sometimes", 3 = "Often" ,4 = "Very Often", 0 = "I don't know")*

- *Avoiding touching your face (1= "Never", 2= "Sometimes", 3= "Often" ,4= "Very Often", 0= "I don't know").*

Lockdown compliance was measured by the following question:

[1] http://www.institut-viavoice.com/

[2] A summary of the project results is presented in Blayac et al. (2022a). For example, Blayac et al. (2022b) tested the effectiveness of a "social comparison nudge" to enhance intention to comply with a future hypothetical lockdown; Blayac et al. (2022c, 2022d) proposed a discrete choice experiment to estimate population preferences in terms of lockdown characteristics; Wen et al. (2022) investigated the preventing role of mindfulness on COVID-19 adverse effects on mood and sleep ; and Wang et al. (2022) investigated the influence of age on time and risk preferences.

"Would you say that you strictly respect the lockdown governmental directive?" (Answer in a 0 to 10 scale, with 0 = "Not at all" and 10 = "strictly").

And mask wearing with the following question:

"During the lockdown, did you wear a mask when you left your home?" (1 = "Never", 2 = "Rarely", 3 = "Often", 4 = "Always").

2.3. Economic preferences and preference measures

We relied on two types of tools to measure economic preferences: EP based on incentivized experimental tasks and SP based on answers to questionnaires. Table 1 provides an overview.

	Stated preferences (SP)	**Elicited preferences (EP)**
Risk preferences	Willingness to take risk general / health / finance (Dohmen et al., 2011)	Task 1: Portfolio choice task (Gneezy and Potters, 1997)
Time preferences	Self reported 0-10 scale	Task 4: Convex Time Budget method (Andreoni and Sprenger, 2012)
Social preferences	Trust in general / professional / family domain: General Social Survey (GSS)	Cooperativeness Task 2: Voluntary contribution to a (linear) public good - PGG Prosociality Task 3: SVO (Murphy et al., 2011)

Table 1: Measures of preferences.

To compare incentivized tasks and self-reported questions, we matched SP and EP for the risk and time dimensions. Regarding social preference, we measured prosociality and cooperativeness with EP and trust with SP. We used the standard GSS question "*Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?*".

The EP block consisted of four different tasks presented in the following order: (1) the portfolio choice task (Gneezy and Potters, 1997), (2) a four-player one-shot linear public good game, (3) the six-item Social Value Orientation (SVO) task (Murphy et al., 2011), (4) a convex time budgeting task (Andreoni and Sprenger, 2012). Participants were told that their performance in only one of these games would be randomly selected at the end of the experiment to determine their payment.

(1) In the portfolio choice task, participants were endowed with 20€ and had to decide how much to invest in a risky asset that paid out either three times the invested amount or zero, with equal probability (the outcome was added to the part of the endowment that was not invested). The amount of money not invested in the portfolio, i.e. the amount which was invested in the risk-free asset, was used as our elicited measure of risk aversion.

(2) In the public good game, participants were endowed with 20€ and had to decide how much to invest in a linear public good with a return rate of 0.5. The amount of the endowment invested in the public good was used as our elicited measure of cooperativeness.

(3) We relied on Murphy et al. (2011) for the six-item Social Value Orientation task (which consists of six dictator-like monetary allocation problems, resolved between the respondent and another randomly selected player) and use the SVO angle as a measure of prosociality (the higher the SVO angle the more prosocial the individual).[3]

(4) Finally, the convex time budget task consisted of two allocation decisions. For each decision, participants had to allocate 40€ between two dates which were one month apart.[4] Each euro allocated to the later date was multiplied by 1.2. We used the average share of the endowment allocated to the earlier date as a measure of discount rate, and took the ratio of the amount allocated to the earlier date for the first and second decisions as a measure of present bias. If individuals are time-consistent, they should allocate the same amount to the earlier date in both decisions. On the other hand, if they are present (future) biased they will allocate a smaller (larger) amount to the earlier date for the second decision than they will for the first; i.e. they will exhibit decreasing (increasing) impatience.

In the SP block (proposed after the EP block) we measured self-reported preferences on a 0-10 Likert scale. First, we asked respondents how willing they were to take risk in *general*, and in the specific domains of their *finance* and *health* (Dohmen et al., 2011). We then asked respondents how *patient* they were (Vischer et al., 2013). Finally, we asked them to state their willingness to trust, based on the General Social Survey (GSS). We used both the standard dichotomous question and an 11-point Likert scale. The questions were repeated to measure *general*, *family* and *professional* trust. An overview of the preference measures, with their mean and standard deviation, is presented in Table 2.

	Variable name	**Definition**	**Incentivized**	**Mean**	**Std**
EP	Risk aversion	Amount invested in the sure asset in the portfolio choice task.	Yes *Payoff*	14.555 *22.17€*	7.619 *14.34€*

[3] In the SVO task, participants were acknowledged that only one of the decisions would be selected for payment.
[4] May, 18th and June, 18th for the first allocation decision, and June, 18th and July, 18th for the second allocation decision

	SVO angle	Social Value Orientation angle (Murphy et al, 2011).	Yes *Payoff*	32.012 *38.23€*	13.290 *13.96€*
	Cooperativeness	Amount of money invested in the group project (public good game).	Yes *Payoff*	5.376 *25.37€*	7.287 *7.29€*
	Discounting	Average share of endowment allocated to the sooner date of the CTB task.	Yes *Payoff (1st date)*	0.348 *13.97€*	0.309 *13.56€*
	Present bias	Ratio of the amount allocated to the sooner date for question 1 (date t and t+1) and question 2 (date t+1 and t+2) of the CTB task.	Yes *Payoff (2nd date)*	2.031 *31.24€*	5.700 *16.27€*
SP	Risk general	Self-reported willingness to take risk in general (0-10)	No	3.931	2.695
	Risk financial	Self-reported willingness to take risk in financial situation (0-10)	No	2.653	2.529
	Risk health	Self-reported willingness to take risk in health situation (0-10)	No	2.335	2.529
	Trust general	Self-reported level of trust in general (0-10)	No	4.457	2.563
	Trust family	Self-reported level of trust toward family members (0-10)	No	7.586	2.756
	Trust professional	Self-reported level of trust toward colleagues (0-10)	No	4.639	2.807
	Patience	Self-reported level of patience (0-10)	No	5.919	2..749

Table 2: Variables' definition, mean and standard deviation.
Note: "gains in the game" are potential, participants were paid (random) for only one of the incentivized tasks.

2.4. Control variables

The survey also collected individual characteristics: gender, age (defined as younger than 25, older than 60, or somewhere in between), household income (household monthly income lower or equal to 2000€, greater or equal to 4000€, or somewhere in between), political opinion (closer to the left, closer to the right, or in the center), potential comorbidities (such as diabetes or obesity) which might increase the vulnerability of the respondent to infection, and living conditions (whether the respondent was living in the same house as a vulnerable or elderly person).

We also controlled for the survey date (between the 4 th and the 16th of May) and extracted the infection rate in the respondents' area. We used the infection rate to construct a dummy indicating whether the respondent was living in an "active circulation area" as classified by the French government (at this time, a particular emphasis was put in the French media on these particular regions). All the control variables are described and summarized in Table 3.

Variable name	**Description**	**Mean (SD)**
Men	= 1 if the respondent is a man	51.09% (0.50)
< 25	=1 if the respondent is 25 or younger	9.53% (0.29)
> 60	= 1 if the respondent is 60 or older	34.14% (0.47)
high income	= 1 if the individual reported a household income greater or equal to 4000 euro per month	29.38% (0.46)
low income	= 1 if the individual reported a household income lower or equal to 2000 euro per month	23.83% (0.43)
Left political opinions	= 1 if the individual reported political opinions closer to the left than the right.	35.70% (0.48)
Right political opinions	= 1 if the individual reported political opinions closer to the right than the left.	29.81% (0.46)
Vulnerable person	= 1 if the respondent is more vulnerable to COVID-19 (diabete, overweight, etc.)	23.92% (0.43)
Liv. with. vuln. pers.	= 1 if the respondent is living with a vulnerable person	20.80% (0.41)
Highly infected area	= 1 if the respondent lived in a region where the COVID-19 incidence rate was classified as high by the French government.	27.47% (0.45)
Survey date	The day of the survey (between May, 4th and 16th)	9.10 (3.09)

Table 3: Definition and distribution of the control variables.

3. Results

Our empirical strategy proceeded as follows. We begin with descriptive statistics in subsection 3.1 by reporting the distributions of the dependent variables, i.e. compliance with the various measures, as well as their correlations. In subsection 3.2 we examine the correlation between SP and EP for each of the preference dimensions (social, time and risk). In subsection 3.3 we present ordered logit regressions with prophylactic measures as dependent variables, in order to assess the explanatory power of each measure of preferences. Subsection 3.4 discusses the goodness of fit of the models, which is compared according to whether or not they included the block of variables measuring stated preferences or the block measuring elicited preferences. Finally, in subsection 3.5 we assess the *out of* *sample* predictive power, based on a computational approach that simultaneously selects the best model and a subset of the most relevant variables.

3.1. Distribution of dependent variables

The distributions of the self-reported levels of compliance with the different prophylactic measures are presented in Figure 1. Most of the population reported a high level of compliance with each prophylactic measure.

Figure 1: Distributions of self-reported compliance with prophylactic measures.

Table 4 shows the pairwise correlation between the different prophylactic measures. All the measures are positively correlated at the 0.1% level (Pearson correlation coefficient r = 0.108 to 0.239), with the exception of social distancing and coughing in sleeves which are correlated only at the 10% level (Pearson correlation coefficient r = 0.051). Despite significant correlations across measures, Cronbach's alpha is equal to 0.495, which is not high enough to construct a single relevant index of compliance which takes all the measures into account. We therefore decided to conduct separate analyses for each prophylactic measure.

	Respecting lockdown	Wearing a mask	Coughing in sleeves	Not touching face	Social distancing
Washing hands	0.239 ***	0.159 ***	0.225 ***	0.216 ***	0.118 ***
Social distancing	0.226 ***	0.211 ***	0.0512·	0.215 ***	
Not touching face	0.193 ***	0.212 ***	0.147 ***		
Coughing in sleeves	0.108 ***	0.125 ***			
Wearing a mask	0.228 ***				

Table 4: Pairwise correlation between self-reported levels of compliance.
*Note: *** indicates Pearson correlation coefficient significant at the 0.1% level. · indicates Pearson correlation coefficient significant at the 10% level.*

3.2. Correlation between elicited and stated measures of preferences

Table 5 shows the correlation between EP and SP measures. With regard to the risk dimension, we observe that risk aversion (measured by the amount invested in a risk-free asset) is positively but weakly correlated with self-reported willingness to take risk in the general (r = 0.138) and in the financial domain (r = 0.178), but not with self-reported willingness to take risk in the health domain (r = 0.046). Our correlations were slightly lower than those of Vieidier et al. (2015), who found r = 0.196 and r = 0.220 for general and financial risk, respectively, in a sample of French students (N = 93).[5] Self-reported willingness to take risk in the different domains were all significantly correlated (r = 0.403 to 0.542). This seems to indicate that SP for risk in a given domain is a good predictor for SP in other domains. With regard to the social dimension, cooperativeness and prosociality were positively but weakly correlated with stated level of trust across the different domains (r = 0.096 to 0.128). Stated measures of trust in the different domains were moderately correlated (r = 0.478 to 0.614). Surprisingly, we found a weak, but significant, negative correlation between cooperativeness and prosociality, as measured by the SVO; this contrasts with the results of Balliet et al. (2009)'s meta-analysis. Finally, with regard to the time dimension, we find no significant correlations (at the 5% level) between present bias, the discount factor and self-reported level of patience.

To summarize, SP and EP measures of preferences were correlated, but these correlations were weak or negligible. This indicates that the two types of measures either capture different dimensions or capture the same dimension with high measurement error. As a comparison, Falk et al. (2016) found a correlation of -0.352 (resp. -0.294) between stated general willingness to take risk (resp. willingness to take risk in the financial domain) and revealed risk aversion. The higher correlations between SP and EP in Falk's study could be explained by the fact that their study was performed on a sample of German students (N=409), rather than on the general population. Students in particular may have a higher level of numeracy and a greater ability to compute and deal with probabilities than the general population, and could thus be less prone to measurement error in the experimental task.

		Risk financial (SP)	Risk health (SP)	Risk general (SP)
Risk dimension	Risk aversion (EP)	-0.178***	-0.046	-0.138***
	Risk general (SP)	0.489***	0.542***	
	Risk health (SP)	0.403***		

[5] Vieider et al. (2015) elicited certainty equivalents (CE) for lotteries, and then calculated risk-premiums for each lottery (defined as the difference between the expected value of the lottery and the CE).The average risk premium of an individual is taken as a measure of risk preference.

Time dimension		Patience (SP)	Present bias (EP)		
	Discount (EP)	-0.058 ·	0.052 ·		
	Present bias (EP)	0.016			
Social dimension		SVO angle (EP)	Trust pro. (SP)	Trust family (SP)	Trust general (SP)
	Cooperativeness (EP)	-0.080 **	0.085 **	0.074 *	0.112 ***
	Trust general (SP)	0.096 *	0.614 ***	0.478 ***	
	Trust family (SP)	0.096 *	0.531 ***		
	Trust pro. (SP)	0.128 ***			

Table 5: Correlations between elicited and stated measures of preferences

*Note: ***, **, *, · indicates Pearson correlation coefficient significant at the 0.1%, 1%, 5%, 10% levels respectively.*

3.3. Explanatory power of measured preferences

We conducted ordered logit regression models with each prophylactic measure as a dependent variable. Explanatory variables included the set of elicited preferences, the set of stated preferences, and the control variables. The results are presented in Table 6 and the Marginal Effect at the Mean of the preference variables are reported in the Appendix (Table A3-A8). We report here only the effects that are significant at the 5% or lower levels. Significance at the 10% level is indicated in the table for informative purposes only: it is neither considered nor discussed as results in the discussion section. All variables measuring preferences have been normalized prior to the regression, therefore the β can be interpreted as normalized effect size as they indicate the change in the link function associated to a change in preference of 1 standard deviation.

	Dependent variable:					
	Washing hands	Not touching face	Coughing in sleeves	Social distancing	Mask wearing	Respecting lockdown
	(1)	(2)	(3)	(4)	(5)	(6)
Elicited Preferences						
Risk aversion	−0.018	0.130	0.056	0.160	0.023	−0.061
	(0.100)	(0.083)	(0.089)	(0.120)	(0.090)	(0.083)
Cooperativeness	0.081	−0.198*	0.025	−0.088	0.039	0.069
	(0.101)	(0.083)	(0.089)	(0.115)	(0.090)	(0.083)
SVO angle	0.002	0.058	−0.045	−0.005	−0.029	−0.007
	(0.069)	(0.057)	(0.061)	(0.080)	(0.061)	(0.057)
Discount rate	0.066	0.013	−0.057	0.070	0.031	0.023
	(0.069)	(0.058)	(0.062)	(0.080)	(0.061)	(0.058)
Present bias	0.071	0.091	0.036	0.044	−0.041	−0.032
	(0.081)	(0.057)	(0.066)	(0.090)	(0.061)	(0.056)
Stated Preferences						
Risk general	0.010	−0.120 ·	0.029	−0.252*	−0.199**	−0.183*
	(0.088)	(0.072)	(0.078)	(0.102)	(0.077)	(0.074)
Risk healt	−0.257**	−0.066	−0.146 ·	−0.079	−0.272***	−0.268***
	(0.080)	(0.069)	(0.075)	(0.091)	(0.072)	(0.068)
Risk financial	0.015	0.036	0.022	0.038	0.054	−0.053
	(0.080)	(0.067)	(0.073)	(0.092)	(0.072)	(0.068)
Trust general	−0.111	−0.301***	−0.057	0.018	−0.102	−0.233**
	(0.091)	(0.076)	(0.080)	(0.104)	(0.080)	(0.076)
Trust family	0.137	0.125 ·	0.038	0.129	0.110	0.154*
	(0.085)	(0.070)	(0.073)	(0.096)	(0.074)	(0.070)
Trust professional	−0.076	0.011	0.031	−0.068	−0.031	−0.048
	(0.094)	(0.078)	(0.082)	(0.112)	(0.083)	(0.077)
Patience	−0.005	0.038 ·	0.031	0.041	0.024	0.073***
	(0.025)	(0.021)	(0.022)	(0.028)	(0.022)	(0.021)
Controls						
Men	−1.006***	−0.378**	−0.567***	−0.115	−0.447***	−0.443***
	(0.147)	(0.117)	(0.126)	(0.162)	(0.126)	(0.116)
< 25	−0.426 ·	−0.267	0.403 ·	−0.938***	0.082	−0.395*
	(0.241)	(0.207)	(0.234)	(0.229)	(0.212)	(0.202)
> 60	−0.499***	0.311*	−0.523***	0.862***	0.632***	0.079
	(0.150)	(0.126)	(0.134)	(0.204)	(0.137)	(0.125)
high income	−0.040	0.431**	−0.024	0.185	0.216	0.247 ·
	(0.165)	(0.135)	(0.148)	(0.193)	(0.143)	(0.135)
low income	−0.074	0.290*	−0.183	−0.044	0.118	−0.045
	(0.176)	(0.145)	(0.154)	(0.197)	(0.158)	(0.145)
left opinions	0.187	−0.021	0.057	0.046	−0.232	0.057
	(0.168)	(0.136)	(0.148)	(0.193)	(0.147)	(0.136)
right opinions	0.113	−0.009	−0.055	−0.030	−0.064	0.026
	(0.172)	(0.142)	(0.151)	(0.198)	(0.152)	(0.141)
vulnerable person	0.129	−0.028	−0.053	0.609**	0.539***	0.285*
	(0.167)	(0.135)	(0.146)	(0.222)	(0.153)	(0.139)
liv. with. vuln. pers.	0.201	0.062	0.048	0.140	0.232	0.136
	(0.175)	(0.141)	(0.151)	(0.208)	(0.155)	(0.140)
highly inf. area	0.233	0.057	−0.051	0.154	0.070	0.308*
	(0.159)	(0.126)	(0.135)	(0.177)	(0.134)	(0.126)
survey date	0.014	−0.009	0.021	−0.029	0.019	−0.015
	(0.022)	(0.018)	(0.019)	(0.025)	(0.019)	(0.018)
Observations	1,141	1,119	1,108	1,141	1,116	1,143
Log Likelihood	−875.910	−1,345.957	−1,152.037	−625.437	−1,137.524	−1,569.537

Table 6: Detailed results of ordered logit regressions.

*Note: • : $p < 0.1$; *: $p < 0.05$; **:$p < 0.01$; ***:$p < 0.001$ All variables measuring preferences have been normalized prior to regression.*

In general, the magnitude of the (normalized) β are higher for the SP than for the EP. We observe that, among the EP set, only cooperativeness was significant at the 5% level and only for compliance with the recommendation to avoid touching one's face [$\beta = -0.198\,; z = -2.378\,; p = 0.017$]. By contrast, the SP set seems to explain individual self-reported attitudes more effectively. Self-reported willingness to take risk in general was negatively correlated with compliance with social distancing [$\beta = -0.252\,; z = -2.474\,; p = 0.013$], mask wearing [$\beta = -0.199; z = -2.578\,; p = 0.010$] and lockdown [$\beta = -0.183; z = -2.488\,; p = 0.013$]. Self-reported willingness to take risk in the health domain was negatively correlated with compliance with hand washing [$\beta = -0.257; z = -3.216; p = 0.001$], mask wearing [$\beta = -0.272; z = -3.757\,; p < 0.001$] and lockdown [$\beta = -0.268\,; z = -3.921\,; p < 0.001$]. The general stated level of trust was negatively correlated with compliance with the recommendation to avoid touching one's face [$\beta = -0.301\,; z = -3.964\,; p < 0.001$] and with lockdown [$\beta = -0.233\,; z = -3.076\,; p = 0.002$], while trust in the family circle was positively correlated with compliance to lockdown [$\beta = 0.154\,; z = 2.193\,; p = 0.028$]. Higher stated patience positively affected lockdown compliance [$\beta = 0.073\,; z = 3.496; p < 0.001$].

When significant, the effects of stated willingness to take risk and patience were in the hypothesized direction. Concerning stated trust, as specified in the introduction, truthfulness may have ambiguous effects on compliance. The positive effect of trust in family and compliance with lockdown, can be attributed to individuals who believe their family is adopting a trustful and compliant attitude and are thus pushed to mimic the compliance norm. On the other hand, negative effects of trust in general on compliance with the recommendation of not touching own face and lockdown can be explained by the fact that believing that other can be trusted could decrease the subjective probability of an infection, thus decreasing the necessity of respecting the recommendation. Concerning the negative relation between revealed cooperativeness and not touching face, the sign of the coefficient was not anticipated, since we assumed that more cooperative individuals should be more compliant. One possible explanation is that individuals who exhibit cooperative behavior in the Public Good Game were conditional contributors and hold different beliefs on others' cooperative attitudes in the PGG and regarding COVID. However, considering the high number of hypotheses tested, it is also likely that this result is a false positive (see section 3.4).

With regard to our control variables, we confirm results from previous studies by finding age and gender differences (e.g. Galasso et al. 2020; Szabo et al. 2020). In particular, men exhibited a lower degree of compliance with most of the prophylactic measures (except social distancing). Younger respondents reported respecting social distancing and lockdown less, while older respondents reported a lower level of compliance with hand washing and the recommendation to cough into one's sleeves, but higher compliance with the recommendation to avoid touching their faces, and with social distancing and mask wearing.

3.4. Model comparison: elicited vs stated preferences.

Lack of robustness due to multiple testing of hypotheses is a potential issue with the regression results reported in Table 6. To address this issue, for each prophylactic measure we compared ordered logit regression models that included different sets of explanatory variables: (a) null models (with no variables) versus models with either EP or SP variables only; (b) models with control variables versus models which combined either control variables and EP or control variables and SP; and (c) models which combined either control variables and EP or control variables and SP, versus a model with all variables (i.e. EP, SP and controls). We assessed the goodness of fit of these models by their log likelihood. We relied on likelihood ratio tests to test whether the introduction of stated or elicited preferences significantly increases the goodness of fit of the models (compared to introducing the same number of uncorrelated variables). Only nested models were compared. The results of these likelihood ratio tests are presented in Table 7.

H0	Variables included in the models	Washing hands	Not touching face	Coughing in sleeves	Social distancing	Mask wearing	Respecting lockdown
EP measures do not increase goodness of fit $\chi^2(5)$	None vs **EP**	2.835 $p = 0.725$	11.478 **$p = 0.043$**	3.019 $p = 0.697$	4.921 $p = 0.426$	1.042 $p = 0.959$	1.374 $p = 0.976$
	Controls vs Controls + **EP**	2.368 $p = 0.796$	10.185 $p = 0.070$	4.078 $p = 0.538$	2.182 $p = 0.824$	1.581 $p = 0.904$	3.553 $p = 0.615$
	Controls + SP vs Controls + SP **+ EP**	1.214 $p = 0.876$	4.511 $p = 0.341$	1.618 $p = 0.801$	4.411 $p = 0.353$	1.154 $p = 0.886$	2.667 $p = 0.615$
SP measures do not increase goodness of fit $\chi^2(7)$	None vs **SP**	31.232 **$p < 0.001$**	38.737 **$p < 0.001$**	7.413 $p = 0.387$	33.140 **$p < 0.001$**	63.844 **$p < 0.001$**	120.435 **$p < 0.001$**
	Controls vs Controls **+ SP**	21.952 **$p = 0.005$**	41.079 **$p < 0.001$**	13.341 $p = 0.101$	23.523 **$p = 0.003$**	52.423 **$p < 0.001$**	102.439 **$p < 0.001$**
	Controls + EP vs Controls + EP **+ SP**	20.798 **$p = 0.004$**	35.406 **$p < 0.001$**	7.645 $p = 0.365$	16.930 **$p = 0.018$**	49.689 **$p < 0.001$**	96.219 **$p < 0.001$**

Table 7: Likelihood ratio tests for models comparison.

Note: All models are ordered logit. For each model comparison, the model under the null hypothesis is nested in the model under the alternative. None indicates models with no explanatory variables (all respondents have the same estimated compliance). Significant p-values in bold.

We found that the introduction of EP variables significantly increased the goodness of fit at the 5% level only for the model of compliance with the recommendation not to touch one's face, and only when compared to a constant model [$\chi^2(5) = 11.478$, $p = 0.043$]. By contrast, adding SP variables significantly increased the goodness of fit of all models, except the models of compliance with the recommendation to cough in one's sleeves and the model of compliance with social distancing (and at the 10% level for the other specifications). Moreover, the "SP" and "Controls + SP" models have higher AIC than the "EP" and "Controls + EP" models for all prophylactic measures, and a higher BIC for all prophylactic measures but the recommendation not to cough in sleeves (see Table 8).

		Washing hands	Not touching face	Coughing in sleeves	Social distancing	Mask wearing	Respecting lockdown
EP	AIC	1890.59	2798.44	2408.14	1380.89	2433.06	3346.45
	BIC	1909.85	2817.59	2427.241	1400.15	2452.19	3353.70
SP	AIC	1866.20	2775.19	2407.74	1356.67	2374.26	3231.39
	BIC	1895.55	2804.38	2436.881	1386.03	2403.44	3248.74
EP + Controls	AIC	1810.62	2765.32	2349.72	1305.80	2362.74	3285.29
	BIC	1885.26	2839.64	2423.89	1380.44	2437.02	3347.96
SP + Controls	AIC	1797.03	2740.42	2346.46	1290.46	2317.89	3192.41
	BIC	1879.75	2822.78	2428.65	1373.17	2400.22	3263.15

Table 8: AIC and BIC comparison between EP and SP models.

From these results, we conclude that our SP variables have stronger overall explanatory power for self-reported compliance with prophylactic measures than the EP variables do.

3.5. Out of sample predictive power

In the last section, we investigated the explanatory power of models containing EP and/or SP variables, to answer the question: "to what extent were these models able to explain actual compliance in a given sample?". However, a policy maker may be more interested in how to *predict* compliance from samples in which individual characteristics and preferences can be measured, than in their *actual* level of compliance. For example, predicting degrees of compliance can be pivotal, before deciding whether to implement or not more stringent prophylactic measures such as lockdown or curfew. To answer this question, one may be tempted to use a model with many variables (like the one presented in Table 6), apply it to a population where prophylactic measures have already been implemented, and then use it to predict compliance in areas where those measures have not yet been implemented.

However, there are two important issues to consider when estimating models with numerous explanatory variables: model mis-specification and over-fitting. Model mis-specification occurs when the shape of the link function $f: X \rightarrow Y$ poorly captures the real relation between the response variable Y (in our case compliance with prophylactic measures) and the explanatory variables X (EP, SP, and individual characteristics). This can happen, for example, if the relation between X and Y is not linear as assumed by the researcher, or if some covariates are correlated and subject to measurement error (Gillen et al., 2019). In these cases, model misspecification is likely to produce over-fitting, particularly when too many variables are introduced in the models. Indeed, if many explanatory variables, not related to the dependent variables are introduced into the model, their associated coefficients will capture random fluctuations of the response variable and attribute part of the true fluctuations to the wrong variables. As a consequence, the estimated model will explain “too much” of the data but will fail to predict unobserved response variables.

An easy way to circumvent these issues is to compare several models’ specifications based on their *out-of-sample* predictive power. Out-of-sample predictive power measures the extent to which a model that has been estimated for a given sample is able to predict the dependent variable for another (unobserved) sample (Clark, 2004).In this section, we rely on a computational methodology proposed by Ellies-Oury et al. (2019), which allows us to simultaneously find (a) the set of explanatory variables and (b) the model (among numerous parametric, non-parametric and semi-parametric candidates) which best predicts (future or unobserved) compliance. This answers two very important methodological questions from the point of view of policy makers: (a) which information to collect and (b) how to treat this information.

The method of Ellies-Oury et al. (2019) is based on a train/test procedure inspired by machine-learning. More precisely, our sample was randomly divided into a training sample (S_{train}) and a testing sample (S_{test}). The training sample was used to estimate the models, and to identify for each model a sub-sample of the variables most strongly linked with *Y*. The testing sample was then used to assess models’ *out of sample* predictive power, by comparing model prediction for the test sample with the response variable. Since the test sample has not been used to estimate the link function, comparing models based on their *out-of-sample* predictive power, is a natural way to prevent over-fitting, and to compare models that can differ both in their number of variables and in the shape of the link function.

For each model, variable selection was made by computing the *variable importance* (VI), defined as the mean square error (MSE) when the *j*-th covariate is randomly permuted, for each covariate *j*:

$$VI_j = \frac{1}{|S_{train}|} \times \sum_{i \in S_{train}} \left(y_i - \widehat{y_i}^{(j)}\right)^2$$

Where y_i the *i*-th observation of the dependent variable, and $\widehat{y_i}^{(j)} = \widehat{f}^{(j)}(X_i)$ is the predicted value when the observations of the *j*-th covariate are randomly permuted and $\widehat{f}^{(j)}$ is the new estimated link function (after permutation). Indeed, the greater the effect of the *j*-th covariate on *Y*, the more negatively the random permutations will affect the quality of predictions, thus increasing the MSE. As suggested by Elies-Oury et al. (2019), we permuted each covariate 50 times, in order to provide robust estimates for *VI*. The variables with the higher *VI* were then selected using a single change point detection in *VI*’s mean and variance, using the algorithm developed by Killick and Eckley (2014).

The model was then reduced by conserving only the subset of variables that had been selected based on the variable importance, and estimated again on the training sample. This estimated model was then used to predict the values in the *test sample*, and we compute the mean squared error in the test sample (MSE_{test}) by comparing the observed values Y_{test} and the predictions $f(X_{test})$:

$$MSE_{test} = \frac{1}{|S_{test}|} \times \sum_{i \in S_{test}} (Y_{test} - f(X_{test}))^2$$

MSE_{test} therefore is a measure of the out-of-sample predictive power of the model, since the values X_{test} are not used to estimate the model, but only to assess its performance.

Ultimately, since the value of the MSE_{test} depends on how the data has been splitted into a training and a testing sample, we repeated the procedure 500 times to obtain robust estimates of MSE_{test}. This procedure was made for each of the prophylactic measures and each of the candidate models (see below).

To summarize, for each candidate model, and each compliance measures, we repeated the following procedure 500 times:

(1) The observations were randomly divided into a *training sample* (80% of the observations) and a *testing sample* (20% of the observations).
(2) The coefficients of the model with all explanatory variables were estimated for the *training sample*.
(3) Each explanatory variable was permuted 50 times, to identify the importance of the variable in predicting compliance. The most important variables were selected using a standard single change point detection algorithm (Killick and Eckley, 2014).
(4) Compliance with the prophylactic measure in the *testing sample* were predicted using the coefficients of the model estimated in (3)
(5) Predicted and reported levels of compliance were compared according to mean squared error, calculated on the *testing sampl*e (MSE_{test}).

The algorithm was implemented in the *modvarsel* package for *R*. We used this methodology to compare several regression models already implemented in the package, and updated the package to integrate the ordered logit model (olm) presented in the previous section. More precisely, we compared: (i) linear regression (linreg), (ii) random forest (rf) (Breiman, 2001), (iii) kernel sliced inverse regression (sir) (Wu, 2008), which is associated with kernel estimation, (iv) principal component regression (pcr), (v) partial least square regression (plsr) (Helland, 1990), (vi) ridge regression (ridge) (Hoerl and Kennard, 1970), and (vii) olm regression:
(i) Because of its simplicity, linear regression is a natural benchmark for any regression model.
(ii) Random forest is one of the most popular algorithms in machine-learning for prediction. It does not make any parametric assumption on the link function but estimates an arbitrarily complex shape, based on the observations. However, when too many variables are fed into it, a random forest algorithm may suffer from the "curse of dimensionality".
(iii) Kernel sliced inverse regressions are semi-parametric models with a nonparametric univariate link function between the response variable and a linear combination of the explanatory variables, estimated using kernel regression.
(iv-v) Principal component regressions and partial least square regressions are both better adapted than linear regressions to situations where there are many variables that may be correlated and subject to measurement errors. These algorithms change the space of the covariates into orthogonal components, and use (some of) these components as explanatory variables in the regression. The difference between the two methods is that plsr uses the information contained in Y to select the components used into the regression.
Finally, (vi) ridge regression is appropriate for the correction of multicollinearity: it introduces an additional constraint on the norm of the estimated parameters to the optimization problem. Ridge estimates are not unbiased, but may have a lower variance if compared to OLS estimates.

Figure 2 shows, for each of those models and for each prophylactic measure, the distribution of the MSE_{test} of 200 reduced models (i.e. variables that did not increase the MSE_{test} were not included in the models). Table 10 shows the mean and standard deviation of the MSE_{test}, as well as the mean and standard deviation of the number of variables included in the reduced models.

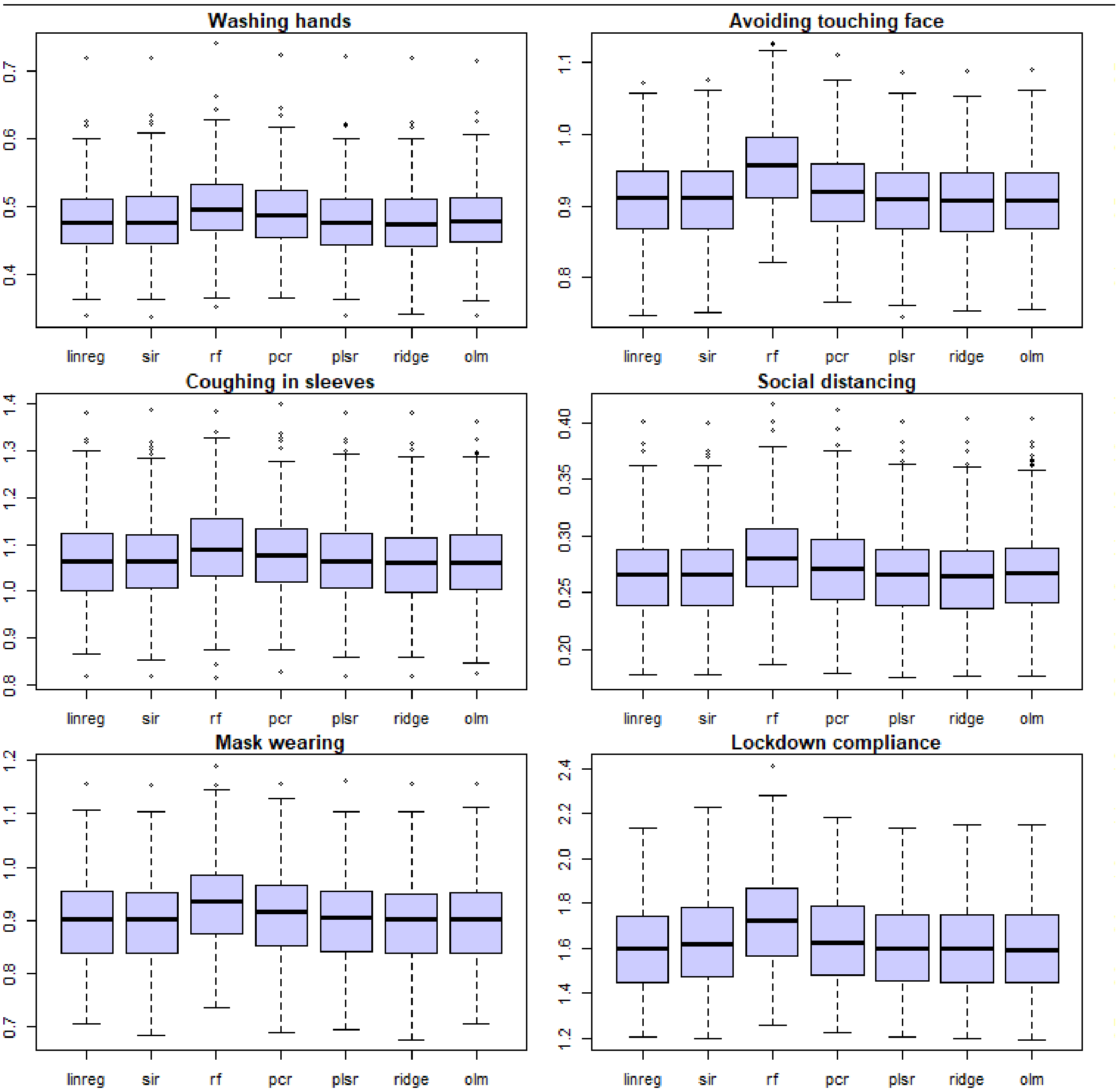

Figure 2: Distribution of mean square errors.

Note: Boxplots of the tested mean square error for 500 replications of the method, after variable selection, for each prophylactic measure and each model type.

		linreg	sir	rf	pcr	plsr	ridge	olm
Washing hands	MSE_{test}	*0.482* (0.049)	*0.484* (0.051)	0.502 (0.051)	0.491 (0.049)	*0.482* (0.049)	**0.481** (0.049)	*0.483* (0.048)
	model size	6.61 (2.15)	5.28 (2.123)	8.04 (5.914)	5.35 (1.952)	6.57 (2.032)	5.61 (1.850)	7.72 (2.410)
Coughing into sleeves	MSE_{test}	*1.055* (0.079)	*1.055* (0.080)	1.087 (0.082)	1.068 (0.079)	*1.054* (0.079)	**1.045** (0.078)	*1.054* (0.079)
	model size	6.7 (3.000)	6.8 (5.980)	7.08 (4.469)	6.16 (2.843)	6.85 (3.209)	5.28 (2.402)	8.72 (3.208)
Not touching face	MSE_{test}	*0.904* (0.057)	*0.905* (0.057)	0.952 (0.059)	0.915 (0.056)	*0.903* (0.056)	**0.901** (0.057)	*0.902* (0.056)
	model size	9.48 (3.088)	7.97 (4.073)	10.88 (6.336)	6.14 (2.651)	9.26 (3.063)	8.5 (2.738)	10.21 (3.816)
Social distancing	MSE_{test}	*0.264* (0.038)	*0.265* (0.038)	0.279 (0.037)	0.271 (0.039)	**0.264** (0.038)	*0.264* (0.038)	*0.266* (0.038)
	model size	6.51 (2.101)	5.29 (3.261)	11.99 (6.902)	7.59 (2.016)	6.6 (2.228)	5.24 (1.93)	8.58 (2.7)
Wearing mask	MSE_{test}	*0.9* (0.083)	*0.899* (0.083)	0.933 (0.083)	0.908 (0.081)	*0.9* (0.083)	**0.896** (0.082)	*0.898* (0.083)
	model size	9.18 (2.013)	7.12 (2.659)	10.4 (5.99)	6.39 (1.68)	8.77 (2.14)	8.03 (2.284)	9.82 (2.493)
Respecting lockdown	MSE_{test}	*1.58* (0.185)	1.611 (0.196)	1.699 (0.186)	1.613 (0.187)	*1.581* (0.185)	**1.577** (0.183)	*1.583* (0.186)
	model size	10.47 (2.148)	8.53 (3.507)	8.74 (5.093)	5.21 (2.133)	10.06 (2.173)	9.33 (2.165)	11.37 (2.592)

Table 10: Summary of models performances.

Note: For each prophylactic measure and each model, the table reports the mean and the standard deviation (in parenthesis) of the mean square error and number of variables after 500 replications of the method (each replication included variable selection). The MSE_{test} of the model with the highest predictive power is indicated in bold. Mean MSE_{test} that are not statistically different from the smallest mean MSE_{test} are in italics.

While the ridge model had a higher predictive power for five out of six measures, its MSE_{test} was not statistically different from the MSE_{test} of the linreg, sir, plsr, and olm models, for all of the prophylactic measures (while the random forest models and the principal component regressions performed significantly worse for all prophylactic measures). Since model size was different among those models, we examine which variables were selected in the most performant models (linreg, sir, plsr, ridge, and olm). For the sake of clarity, we only report, for each prophylactic measure, the EP and SP variables that were selected in more than half of the models for at least one regression method. Detailed results, indicating the selection rate are presented in appendix (Table A10).

Concerning EP variables: none of them was selected in most of the models predicting *Coughing in sleeves*, *Social distancing*, *Mask wearing* and *Respecting lockdown*, regardless of the regression method (linreg, sir, plsr, ridge, and olm).

- *Cooperativeness* was selected in more than a half of the models predicting *Washing hands,* for the linreg, plsr and olm models (between 51% and 52%), and in more than a half of the models predicting *Not touching face* (all regression methods between 50% and 90%).
- *Risk aversion* was selected in more than half of the models predicting *Not touching face,* for ridge (51%) and olm (73%) regressions.
- *Present bias* was selected in more than half of the models predicting *Not touching face*, for the ridge, olm, plsr, linreg, and plsr regressions (between 52% and 59%).
- *SVO angle,* and *Discount rate* were not selected in most of the models, for all prophylactic measures and for all regression methods.

The selection rate of SP variables is in general higher. For all prophylactic measures and for all regression methods, most of the models selected at least one SP variable.

- *Risk general* was selected in most of the models predicting *Coughing in sleeves,* for the ridge, linreg, plsr, and olm regressions (between 51% and 68%), in most of the models predicting *Social distancing* (all regression methods between 57% and 94%), *Mask wearing* (all regression methods between 85% and 94%) and *Respecting lockdown* (all regression methods between 78% and 98%).
- *Risk health* was selected in almost all the models predicting *Washing hands* (all regression methods between 93% and 99%), *Mask wearing* (all regression methods between 85% and 100%), and *Respecting lockdown* (all regression methods between 99% and 100%) and most of the models predicting *Coughing in sleeves* (all regression methods between 66% and 89%).
- *Trust general* was selected in almost all the models predicting *Not touching face* (all regression methods between 98% and 100%), and *Respecting lockdown* (all regression methods between 83% and 98%). It was also selected in most of the models predicting *Coughing in sleeves* (all regression methods except sir between 55% and 72%), and in most of the olm regressions (58%) predicting *Washing hands.*
- *Trust family* was selected in almost all the models predicting *Respecting lockdown* (all regression methods between 90% and 98%), in most of the models predicting *Washing hands*, for the linreg, plsr, and olm regressions (from 53 to 79%), and in most of the olm regressions predicting *Not touching face* (69%), *Social distancing* (72%), and *Mask wearing* (79%).
- *Patience* was selected in most of the models predicting *Not touching face* (all regression methods except sir between 60% and 70%), in 58% of the olm regressions predicting *Social distancing*, and in most of the models predicting *Respecting lockdown* (all regression methods between 52% and 95%).
- *Risk financial* and *Trust pro* were not selected in most of the models, regardless of the prophylactic measure and of the regression method considered.

4. Discussion

We studied the predictive power of experimentally elicited standard measures of preferences (EP) and stated preferences (SP) for self-reported compliance to COVID-19 prophylactic recommendations and restrictions. Our main finding is that EP our elicited measures, based on incentivized experimental tasks, are poor predictors in contrast to standard measures based on responses to hypothetical questions. This finding applies to both self-centered preferences, i.e. risk and time preferences, and to social preferences, i.e. trustfulness, cooperativeness and other-regardingness. This study therefore contributes to the growing body of literature in which the concordance between SP methods and revealed preferences methods based on incentivized tasks is investigated. We documented a discrepancy between measures in the case of self-reported compliance with COVID-19 restrictive measures, based on a large sample which was representative of the metropolitan French adult population in terms of age, gender, and area of residence. Our results challenge the validity of both experimental methods and declarative methods in predicting stated real-world behavior. Ultimately, our findings point towards the need to understand the reasons for the gap between the two types of methods (Hertwig et al., 2019). In this section, we provide an extensive discussion about this discrepancy with reference to the two types of preferences most studied in the literature: risk preferences and trust.

Risk preferences have so far received the most attention in the literature. Several key papers (Dohmen et al., 2011; Vieider et al., 2015; Falk et al., 2018) have established a strong correlation between survey measures and incentivized measures of risk preferences based on large samples: the German national panel (N = 22 000 for the survey measure and N = 450 for the experimental measure) for Dohmen et al. (2011), a 30-nations sample (N = 2939) for Vieider et al. (2015) and a world-wide (76 countries) sample (N = 80 000 for the survey and N = 409 for the experimental measure) for Falk et al. (2018). In addition, Anderson and Mellor (2008) found that elicited risk preferences (using the Holt and Laury (2002) MPL method) in a large US (non-representative) sample (N = 1094) are associated with smoking, drinking and obesity. In contrast, using a representative sample of the UK population, Galizzi et al. (2016) found mixed evidence supporting the external validity of experimentally measured risk attitudes: these measures did not predict smoking status, junk food consumption, regular savings and savings' time horizons. Similarly, Charness et al. (2020) found that neither stated risk preferences (as in Dohmen et al., 2011) nor elicited preferences (as in Gneezy and Potters,1997, and in our study) are able to predict risky decisions in practice (e.g. amount of savings, share of risky investments, holding financial insurance and health insurance) for a representative sample of the Dutch population. In another example, Frey et al. (2017) found weak support in favor of correlation between experimental methods and stated preference methods (N = 1500). Taken together, these findings suggest that evidence for the predictive power of preferences elicited in the lab for outside-lab behavior is mixed.

A related issue is the plurality of experimental methods used to elicit risk preferences, and the absence of consensus about the most appropriate method (Charness et al., 2013). For this reason, the correlation, or absence of correlation, between elicited and stated preferences has been established for only a few of these methods (e.g., the CE method). Several papers have pointed out the low correlation between different measures of risk preferences (Anderson and Mellor, 2008; Reynaud and Couture, 2012; Deck et al., 2013; Dullek and Fell, 2013). Recently, the *risk elicitation puzzle* introduced in Pedroni et al. (2017) cast serious doubts on the relevance of experimentally elicited risk preferences. The authors compared the outcome of six risk preference elicitation methods using a within-subject design, and found considerable variation across methods. Their conclusion suggests a possible inconsistency in the measurement of risk preferences, contradicting procedural invariance. Also using a within-subject design, Holzmeister and Stefan (2021) replicated Pedroni et al. (2017)'s findings with other elicitation methods. However, they adopted a more cautious posture in their conclusion: elicited preference variation across methods could either be a violation of procedural invariance, a violation of the stability axiom of risk preferences, or both. According to Holzmeister and Stefan (2021), concluding that elicited preferences are inconsistent requires "the usage of different risk preference elicitation methods to compare the elicited preferences, which (implicitly) assumes procedural invariance — and vice versa". In short, one needs a theoretical framework before drawing a sharp conclusion. Interestingly, Holzmeister and Stefan (2021) found that their participants were aware of the risk they took in each elicitation method, i.e. risks were taken deliberately depending on the method; this suggests that participants understood the method and reacted to it (see Crosetto and Fillipin, 2016, for a similar conclusion). This observation also supports the conclusion of Vieider et al.

(2015) about the existence of a common risk preference factor that is independent of the risk domain and of the elicitation method. Similarly, Frey et al. (2017) suggest the existence of a common generalized risk factor *R* (similar to the *g-factor* for cognitive ability) that is supposedly independent of the elicitation method.

As in the case of risk-preferences, there is also mixed evidence about the correlation between survey measures and behavioral measures of social preferences, such as trust or cooperativeness. Levitt and List (2007) analyzed the generalizability of lab findings about social preferences to real world settings. Recently, Galizzi and Navarro-Martinez (2019) carried out a systematic review and found mixed evidence that EP in dictator, ultimatum and public good games correlated with charitable behaviors observed in the field or with GSS survey questions. They conducted their own large-scale field-lab experiment and showed that the behavior exhibited in incentivized social dilemmas is neither a good predictor of self-reported measures of social preferences nor of field behavior. In terms of cooperativeness, Reindl et al. (2019) observed that cooperativeness in the lab translates to a field setting: subjects who are more cooperative in the lab are also more cooperative in the field, and vice-versa.

Trust is one of the most studied dimensions of social preferences. The standard SP measure, replicated in our survey, is the GSS binary question; while the standard EP measure is based on the investment game (Berg et al., 1995) or its reduced version, the trust game (Kreps, 1997). Several papers found weak correlation between the SP and the EP measures for trust: these included Gleaser et al. (2000), Lazzarini et al. (2005), Ashraf et al. (2006), Bellemare and Kroeger (2007), and Emrish et al. (2009). However, Fehr et al. (2005) found strong positive correlation between elicited and stated trust. Nevertheless, the bulk of available evidence so far weighs against a positive correlation. Again, there are several reasons. On the experimental side, the games that are used to elicit trust are probably inappropriate, because of the many confounding factors (Alós-Ferrer and Farolfi, 2019) including risk aversion (Eckel and Grossman, 2004; Fetchenhauer and Dunning, 2012), betrayal aversion (Bohnet and Zeckhauser, 2004) and other-regarding preferences (Cox, 2004). On the SP side, there are many methodological issues around question framing and measurement (Evans and Revelle, 2008; Alós-Ferrer and Farolfi, 2019). It is fair to say that SP and EP measures of trust probably capture different facets (Alós-Ferrer and Farolfi, 2019) which are weakly related or even unrelated, and which are highly culturally-dependent (Yamagishi and Yamagishi, 1994).

Hertwig et al. (2019) point out the need to understand the methodological gap between behavioral-based methods and stated preferences methods, particularly with regard to their predictive power for issues related to health and well-being. The above discussion suggests that SP might be more appropriate for predicting stated behavior, such as stated compliance behavior, and that EP might be more appropriate for predicting revealed behavior, such as effective compliance. However, evidence for such parallelism is scarce. In the case of risk preferences, Charness et al. (2020) found that EP are good predictors for risk-taking behavior in the lab, but not for such behavior outside the lab. In the case of social preferences, Krupka and Weber (2013) found that experimentally elicited social norms are good predictors for sharing behavior in the lab. Obviously, more evidence is needed before one can recommend the appropriate method, which also depends on the aims of the researcher.

An alternative path is simply to escape the previous debate by relying on experimentally validated survey measures (Dohmen et al., 2011, Vieider et al., 2015, Falk et al., 2016, 2018). EP are closest to the economists' notion of preferences, because they are grounded in the axiom of revealed preferences: this provides a unique connection between consumer theory and the real world. However, eliciting EP properly requires high control and precision. From a practical viewpoint, SP are easier (and costless) to implement, but have two major drawbacks: they are not incentive-compatible, and they offer a plethora of possibilities among which it is almost impossible to choose without a proper theoretical framework. Experimentally validated survey measures offer an attractive compromise by providing an indirect link between SP and the axiom of revealed preference through the mediation of EP, which are incentive compatible.

5. Conclusion

The key strategies for the eradication of the COVID-19 pandemic pertain to limiting the spread of the disease. Prophylactic measures (barrier gestures, mask wearing, lockdown and curfews) were very much in demand at the outbreak of the pandemic, and were maintained or strengthened throughout the crisis, especially because it is known that the vaccine is not sufficient for the curbing of the viral circulation (although very efficient to avoid severe cases). The effectiveness of such measures, however, depends to a great extent on the population's compliance. We studied the determinants of compliance with prophylactic measures (barrier gestures, mask wearing and lockdown) for a representative sample of the metropolitan French population (N = 1154). Two types of determinants were considered: observables, including age, gender, education level, income, house type, and household composition, and unobservables, including economic and social preferences.

In contrast to observable determinants, getting to know people's preferences requires an observational tool. The two most commonly used tools are the stated preferences (SP) and the experimentally elicited preferences (EP) methods. From a theoretical point of view, EP methods rely on the axiom of revealed preference. They are therefore theoretically superior, as they are incentive compatible; in contrast, SP methods suffer from hypothetical bias and "cheap talk". Our web-based surveys and lab-in-the field experiments combined the two methods to determine whether some measures were good predictors of reported compliance. We repeated this exercise for risk-preferences, time-preferences and social preferences.

We report three main findings: (i) the observable determinants of compliance, identified by our study, confirm those found in previous studies about COVID-19 (e.g. Galasso et al. 2020; Szabo et al. 2020), (ii) EP are poor predictors of stated compliance, and (iii) SP are reasonably appropriate predictors of stated compliance. These results are the outcome of a three-step empirical strategy. First, we analyzed the correlation between SP and EP for each preference-dimension and found weak or insignificant correlation, suggesting that the two methods capture different aspects of the same underlying preferences. Second, we estimated compliance for each prophylactic measure, based on ordered logit models, and contrasted the respective explanatory power of SP and EP measures in terms of goodness-of-fit, controlling for the various observable determinants. Our set of SP predictors clearly outperformed our set of EP predictors for each prophylactic measure, with the exception of "avoiding touching one's face" for which both set of predictors performed at an equal level. Third, based on a computational methodology, we performed an out-of-sample predictive power analysis to assess the robustness of our main finding. The results of this exercise confirm the higher predictive power of our SP measures, and allow us to identify the main behavioral determinants of compliance across alternative models.

From an economist's point of view, our results are somewhat disappointing, and puzzling to say the least. They tend to lead towards the rejection of some theoretically grounded EP measures in favor of SP measures, which are prone to hypothetical bias. On the other hand, the knowledge that some SP measures could provide better predictors than some EP measures is useful for policy makers, as these methods are easy and fast to implement. In section 4, we provided an extensive discussion about the relevance of the two methods and the current trend of relying on experimentally validated SP methods (Dohmen et al., 2011, Vieider et al., 2015, Falk et al., 2016, 2018). The route we describe allows for mitigation of the lack of correlation between SP and EP by building correlation explicitly into the methodology.

An important limitation of our study is the measurement of respondents' compliance, that we were not able to directly observe. Instead, our survey relied on stated compliance. It is therefore likely that SP are better predictors, simply because both stated behavior and stated preferences pertain to a common cognitive process that converts actual preferences and behaviors into statements made by the same respondents. Even if most people have a preference for telling the truth (Abeler et al., 2019), they also honestly hold biased motivated memories about their own behaviors (Saucet and Villeval, 2019). This limit is strengthened by the fact that both stated measures of preferences and stated compliance were collected in the same survey. Indeed, if the respondents had a given mindset while answering the survey or wanted to look in a particular way to the experimenters, this could increase the correlation between SP and stated compliance. Moreover, the stated preferences and compliance have been collected in a particular context of the COVID-19 pandemic. In this context, one should consider that government have communicated through media about "avoiding risk", "being patient", "protecting other", in relation with respecting the COVID-19 recommendations, thus increasing the salience of the issues of risk and patience during the crisis, which could also explain the higher correlation between SP measures and compliance relative to rather neutral economic monetary games.[6]

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, Vol. 87, n° 4, p. 1115–1153.

Alós-Ferrer, C., & Farolfi, F. (2019), Trust Games and Beyond. *Frontiers in Neuroscience*, Vol. 13, n° September, p. 1–14.

Andreoni, J., & Sprenger, C. (2012). Estimating time preferences from convex budgets. *American Economic Review*, 102(7), 3333-56.

Arslan, R. C., Brümmer, M., Dohmen, T., Drewelies, J., Hertwig, R., and Wagner, G. G. (2020). How people know their risk preference. *Scientific Reports*, Nature Publishing Group UK, Vol. 10, n° 1, p. 1–14.

Ashraf, N., Bohnet, I., and Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, Vol. 9, n° 3, p. 193–208.

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, *12*(4), 533-547.

Barrios, J. M., Benmelech, E., Hochberg, Y. V., Sapienza, P., & Zingales, L. (2021). Civic capital and social distancing during the Covid-19 pandemic. *Journal of Public Economics*, *193*, 104310.

[6] We thank two anonymous reviewers for pointing out these issues.

Bellemare, C., & Kröger, S. (2007), On representative social capital. *European Economic Review*, Vol. 51, n° 1, p. 183–202

Bertin, P., Nera, K., & Delouvée, S. (2020). Conspiracy Beliefs, Rejection of Vaccination, and Support for hydroxychloroquine: A Conceptual Replication-Extension in the COVID-19 Pandemic Context. *Frontiers in Psychology*, 11, 2471.

Blayac, T., Dubois, D., Duchene, S., Nguyen-Van, P., Rafaï, I., Ventelou, B., & Willinger, M. (2022a). Les attendus d'une approche d'économie comportementale pour les décisions individuelles face à la pandémie de COVID-19: succès et déceptions. *médecine/sciences*, *38*(6-7), 594-599.

Blayac, T., Dubois, D., Duchêne, S., Nguyen Van, P., Rafai, I., Ventelou, B., & Willinger, M. (2022b). Nudging for lockdown. Behavioural insights from an online experiment. *Social Psychology*, https://doi.org/10.1027/1864-9335/a000483

Blayac, T., Dubois, D., Duchêne, S., Nguyen-Van, P., Ventelou, B., & Willinger, M. (2022c). Designing acceptable anti-COVID-19 policies by taking into account individuals' preferences: evidence from a Discrete Choice Experiment. *Working paper*.

Blayac, T., Dubois, D., Duchêne, S., Nguyen-Van, P., Ventelou, B., & Willinger, M. (2022d). Population preferences for inclusive COVID-19 policy responses. *The Lancet Public Health*, 6 (1)

Campos-Mercade, P., Meier, A. N., Schneider, F. H., & Wengström, E. (2021). Prosociality predicts health behaviors during the COVID-19 pandemic. *Journal of Public Economics*, 104367.

Charness, G., Garcia, T., Offerman, T., & Villeval, M. C. (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory?. *Journal of Risk and Uncertainty*, *60*(2), 99-123

Charness, G., & Gneezy, U. (2010), Portfolio choice and risk attitudes: An experiment. *Economic Inquiry*, Vol. 48, n° 1, p. 133–146.

Clark, T. E. (2004). Can out-of-sample forecast comparisons help prevent over-fitting?. *Journal of Forecasting*, *23*(2), 115-139.

Chuang, Y., and Schechter, L. (2015), Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics*, Vol. 117, n° November 2015, p. 151–170.

de Corte, K., Cairns, J., and Grieve, R. (2021), Stated versus revealed preferences: An approach to reduce bias. *Health Economics*, Vol. 30, n° 5, p. 1095–1123.

Craig, A. C., Garbarino, E., Heger, S. A., and Slonim, R. (2017), Waiting to give: Stated and revealed preferences . *Management Science*, Vol. 63, n° 11, p. 3672–3690.

Crosetto, P., and Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics*, Springer US, Vol. 19, n° 3, p. 613–641.

de-Magistris, T., Gracia, A. and Nayga, R.M., Jr. (2013), On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American Journal of Agricultural Economics*, 95: 1136-1154.

Deck, C., Lee, J., Reyes, J. A., & Rosen, C. C. (2013). A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes. *Journal of Economic Behavior & Organization*, 87, 1–24.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, Vol. 9, n° 3, p. 522–550.

Durante, R., Guiso, L., & Gulino, G. (2021). Asocial capital: Civic culture and social distancing during COVID-19. *Journal of Public Economics*, *194*, 104342.

Ellies-Oury, M. P., Chavent, M., Conanec, A., Bonnet, M., Picard, B., & Saracco, J. (2019). Statistical model choice including variable selection based on variable importance: A relevant way for biomarkers selection to predict meat tenderness. *Scientific Reports*, 9(1), 1-12.

Falk, A., Becker, A., Dohmen, T., and Huffman, D. (2016), The preference survey module: a validated instrument for measuring risk, time, and social preferences. *IZA Discussion Papers*, No. 9674

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*, Vol. 133, n° 4, p. 1645–1692.

Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J., and Wagner, G. G. (2005). A nation-wide laboratory: examining trust and trustworthiness by integrating behavioral experiments into representative surveys. *SSRN Electronic Journal*, n° 715.

Fetchenhauer, D., and Dunning, D. (2012). Betrayal aversion versus principled trustfulness – how to explain risk avoidance and risky choices in trust games. *Journal of Economic Behavior & Organization*, 81, 534–541. doi: 10.1016/j.jebo.2011.07.017

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative ? Evidence from a public goods experiment. *Economics Letters*, Vol. 71, n° 3, p. 397–404.

Fischbacher, U., and Gaechter, S. (2010). Heterogeneous social preferences and the dynamics of free riding in public goods experiments. *American Economic Review*, Vol. 100, n° 1, 541–556.

Fossen, F. M., and Glocker, D. (2017), Stated and revealed heterogeneous risk preferences in educational choice. *European Economic Review*, Elsevier B.V., Vol. 97, p. 1–25.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science advances*, *3*(10), e1701381.

Gadarian, S. K., Goodman, S. W., & Pepinsky, T. B. (2021). Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic. *Plos one*, *16*(4), e0249596.

Galasso, V., Pons, V., Profeta, P., Becher, M., Brouard, S., & Foucault, M. (2020). Gender differences in COVID-19 attitudes and behavior: Panel evidence from eight countries. *Proceedings of the National Academy of Sciences*, *117*(44), 27285-27291.

Galizzi, M. M., Machado, S. R., & Miniaci, R. (2016). Temporal stability, cross-validity, and external validity of risk preferences measures: experimental evidence from a UK representative sample. *London School of Economics and Political Science, LSE Library* (No. 67554).

Galizzi, M. M., & Navarro-Martinez, D. (2019). On the external validity of social preference games: a systematic lab-field study. *Management Science*, 65(3), 976-1002.

Geoffard, P. Y., & Philipson, T. (1996). Rational epidemics and their public control. *International economic review*, 37(3), 603-624

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2), 631-645.

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4), 1826-1863.

Hertwig, R., Wulff, D. U., and Mata, R. (2019). Three gaps and what they may mean for risk preference. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, Vol. 374, n° 1766.

Holzmeister, F., and Stefan, M. (2021). The risk elicitation puzzle revisited: Across-methods (in)consistency?. *Experimental Economics*, 24, 593-616.

Howard, G., Roe, B. E., Nisbet, E. C., and Martin, J. F. (2017), Hypothetical bias mitigation techniques in choice experiments: do cheap talk and honesty priming effects fade with repeated choices?, *Journal of the Association of Environmental and Resource Economists*, Vol. 4, n° 2, p. 543–573.

Jacquemet, N., Joule, R.-V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management*, *65*(1), 110-132.

Keser, C., and Winden, F. Van. (2000). Conditional Cooperation and Voluntary Conditional Cooperation Contributions to Public Goods. *The Scandinavian Journal of Economics*, Vol. 102, n° 1, p. 23–39.

Killick, R., & Eckley, I. (2014). Changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, *58*(3), 1-19.

Kreps, D. M. (1990). “Corporate culture and economic theory”, in Perspectives on Positive Political Economy, eds J. E. Alt and K. A. Shepsle (Cambridge: Cambridge University Press), 90–143.

Krupka, E. L., and Weber, R. A. (2013), Identifying social norms using coordination games: Why does dictator game sharing vary?. *Journal of the European Economic Association*, Vol. 11, n° 3, p. 495–524.

Kuiper, M.E., de Bruijn, A.L., Folmer, C.R., Olthuis, E., Brownlee, M., Kooistra, E.B., Fine, A., van Rooij, B., 2020. “The intelligent lockdown: compliance with COVID-19 mitigation measures in the Netherlands.” *Amsterdam Law School Research Paper. No. 2020-20*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3598215

Labiris G, Panagiotopoulou EK, Perente A, Chatzimichael E, Fotiadis I, et al. (2021) “Determinants of compliance to the facemask directive in Greece: A population study.” *PLOS ONE* 16(3): e0248929. https://doi.org/10.1371/journal.pone.0248929

Levitt S, List J. (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J. Econom. Perspect*. 21(2):153–174.

Lim, K. H., Grebitus, C., Hu, W., & Nagya, R. (2015). “Can the ten commandments mitigate hypothetical bias?” University of Arkansas.

List, J. A., and Gallet, C. A. (2001), What experimental protocol influences disparities between actual and hypothetical stated values?, *Environmental and Resource Economics*, Vol. 20, n° 3, p. 241–254.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, Vol. 45, n° 6, p. 633–644.

Müller, S., & Rau, H. A. (2021). Economic preferences and compliance in the social stress test of the COVID-19 crisis. *Journal of Public Economics*, *194*, 104322.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771-781.

Nivette, A., Ribeaud, D., Murray, A., Steinhoff, A., Bechtiger, L., Hepp, U., ... & Eisner, M. (2020). Non-compliance with COVID-19-related public health measures among young adults in Switzerland: Insights from a longitudinal cohort study. *Social Science & Medicine*, 268, 113370.

Painter, M., & Qiu, T. (2021). Political beliefs affect compliance with government mandates. *Journal of Economic Behavior & Organization*, *185*, 688-701.

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour*, 1–26.

Plohl, N., & Musil, B. (2021). Modeling compliance with COVID-19 prevention guidelines: The critical role of trust in science. *Psychology, Health & Medicine*, 26(1), 1-12.

Reindl, I., Hoffmann, R., & Kittel, B. (2019). Let the others do the job: Comparing public good contribution behavior in the lab and in the field. *Journal of Behavioral and Experimental Economics*, 81, 73-83.

Reynaud, A., & Couture, S. (2012). Stability of risk preference measures: Results from a field experiment on French farmers. *Theory and Decision*, Vol. 73, n° 2, p. 203–221.

Sailer, M., Stadler, M., Botes, E. Fischer, F. & Greiff, S. (2022). Science knowledge and trust in medicine affect individuals’ behavior in pandemic crises. *European Journal of Psychology of Education*, 37(1), 279-292

Saucet, C., & Villeval, M. C. (2019). Motivated memory in dictator games. *Games and Economic Behavior*, 117, 250-275.

Sheth, K. & Wright, G.C. (2020) The usual suspects: do risk tolerance, altruism, and health predict the response to COVID-19?. *Review of Economics of the Household,* 18, 1041–1052. https://doi.org/10.1007/s11150-020-09515-w

Szabo, A., Ábel, K., & Boros, S. (2020). Attitudes toward COVID-19 and stress levels in Hungary: Effects of age, perceived health status, and gender. *Psychological Trauma: Theory, Research, Practice, and Policy*, *12*(6), 572.

Vieider, F. M., Lefebvre, M., Bouchouicha, R., Chmura, T., Hakimov, R., Krawczyk, M., & Martinsson, P. (2015). Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries. *Journal of the European Economic Association*, Vol. 13, n° 3, p. 421–452.

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., & Wagner, G. G. (2013), Validating an ultra-short survey measure of patience. *Economics Letters*, Elsevier B.V., Vol. 120, n° 2, p. 142–145.

Wang, Z., Rafaï, I., & Willinger, M. (2022). How age affects risk and time preferences: Evidence from a representative sample. *Available at SSRN 4031930*. http://dx.doi.org/10.2139/ssrn.4031930

Wen, X., Rafaï, I., Duchêne, S., & Willinger, M. (2022). Did Mindful People Do Better during the COVID-19 Pandemic? Mindfulness Is Associated with Well-Being and Compliance with Prophylactic Measures. *International Journal of Environmental Research and Public Health*, *19*(9), 5051.

Wright, A. L., Sonin, K., Driscoll, J., & Wilson, J. (2020). Poverty and economic dislocation reduce compliance with covid-19 shelter-in-place protocols. *Journal of Economic Behavior & Organization*, 180, 544-554

Wu, H. M. (2008). Kernel sliced inverse regression with applications to classification. *Journal of Computational and Graphical Statistics*, *17*(3), 590-610.

Yamagishi, T., and Yamagishi, M. (1994), Trust and commitment in the United States and Japan, *Motivation and Emotion*, Vol. 18, n° 2, p. 130–166.

Zajenowski, M., Jonason, P.K., Leniarska, M., Kozakiewicz, Z., (2020). Who complies with the restrictions to reduce the spread of COVID-19? personality and perceptions of the COVID-19 situation. *Personality and individual differences*, 166, 110199.

Zettler, I., Schild, C., Lilleholt, L., Kroencke, L., Utesch, T., Moshagen, M., ... & Geukes, K. (2022). The role of personality in COVID-19-related perceptions, evaluations, and behaviors: Findings across five samples, nine traits, and 17 criteria. *Social Psychological and Personality Science*, 13(1), 299-310.

Appendices

Appendix 1: Representativeness of our sample

Living area	*Our Sample*	*%Sample*	*Metrop. France*	*%metrop. France*
Auvergne-Rhône-Alpes	*140*	*12,32%*	*8.032.377*	*12,38%*
Bourgogne-Franche-Comté,	*62*	*5,46%*	*2.783.039*	*4,29%*
Bretagne	*57*	*5,02%*	*3340379*	*5,15%*
Centre-Val-de-Loire	*42*	*3,70%*	*2.559.073*	*3,94%*
Corse	*7*	*0,62%*	*344.679*	*0,53%*
Grand Est	*115*	*10,12%*	*5.511.747*	*8,49%*
Hauts-de-France	*90*	*7,92%*	*5.962.662*	*9,19%*
Ile-de-France	*202*	*17,78%*	*12.278.210*	*18,92%*
Normandie	*52*	*4,58%*	*3.303.500*	*5,09%*
Nouvelle-Aquitaine	*96*	*8,45%*	*5.999.982*	*9,25%*
Occitanie	*120*	*10,56%*	*5.924.858*	*9,13%*
Pays de la Loire	*66*	*5,81%*	*3.801.797*	*5,86%*
Provence-Alpes-Côte d'Azur	*87*	*7,66%*	*5.055.651*	*7,79%*
Total	*1136*	*100%*	*64.897.954*	*100%*

Table A1: Representativeness of the sample in terms of living area.
Note: metropolitan French population data was acquired from the French National Institute of Statistics and Economic Studies (INSEE), January 2020. In our sample, living area information was missing for 18 participants.

Âge	Gender	Our sample	%Sample	France	%France
[18-25]	Women	57	4,99%	3.020.365	5,74%
	Men	48	4,20%	3.109.925	5,91%
	Total [18-25]	105	9,19%	6.130.290	11,65%
]25-60[	Women	352	30,80%	14.625.305	27,78%
	Men	294	25,72%	14.036.119	26,67%
	Total]25-60[	646	56,52%	28.661.424	54,45%
[60,+∞[	Women	150	13,12%	9.967.825	18,94%
	Men	242	21,17%	7.882.415	14,97%
	Total [60,+∞[	392	34,29%	17.850.240	33,91%
Total	Women	559	48,91%	27.613.495	52.46%
	Men	584	51,09%	25.028.459	47.54%
	Total	1143	100%	52.641.954	100%

Table A2: Representativeness of the sample in terms of gender and age

Appendix 2: Detailed regression results.

	1 = "Never"	*2 = "Sometimes"*	*3 = "Often"*	*4 = "Very Often"*
Risk aversion	0.0001 (0.0016)	0.0005 (0.0059)	0.0009 (0.0111)	-0.0016 (0.0187)
Cooperativeness	-0.0015 (0.0016)	-0.0057 (0.0059)	-0.0107 (0.0111)	0.0180 (0.0186)
SVO angle	-0.0002 (0.0011)	-0.0006 (0.0041)	-0.0011 (0.0077)	0.0018 (0.0129)
Discount rate	-0.0013 (0.0012)	-0.0047 (0.0042)	-0.0088 (0.0078)	0.0148 (0.0131)
Present bias	-0.0009 (0.0012)	-0.0034 (0.0046)	-0.0064 (0.0085)	0.0107 (0.0143)
Risk general	0.0000 (0.0014)	-0.0001 (0.0053)	-0.0002 (0.0099)	0.0003 (0.0167)
Risk health	0.0043*** (0.0016)	0.0158** (0.0051)	0.0295** (0.0093)	-0.0496** (0.0153)
Risk financial	-0.0003 (0.0013)	-0.0009 (0.0048)	-0.0018 (0.0090)	0.0030 (0.0151)
Trust general	0.0015 (0.0015)	0.0056 (0.0054)	0.0105 (0.0101)	-0.0177 (0.0170)
Trust family	-0.0021 (0.0014)	-0.0080 (0.0051)	-0.0149 (0.0095)	0.0250 (0.0159)
Trust professional	0.0014 (0.0015)	0.0051 (0.0056)	0.0095 (0.0104)	-0.0160 (0.0175)
Patience	0.0001 (0.0004)	0.0006 (0.0015)	0.0010 (0.0028)	-0.0017 (0.0048)

Table A3 –Marginal Effect at the mean (MEM) for Washing hands

*Note: · : $p < 0.1$; *: $p < 0.05$; **:$p < 0.01$; ***:$p < 0.001$ All variables measuring preferences have been normalized prior to regression.*

	1 = “Never”	*2 = “Sometimes”*	*3 = “Often”*	*4 = “Very Often”*
Risk aversion	-0.0073 (0.0047)	-0.025 (0.0131)	0.0022 (0.0017)	0.0256 (0.0163)
Cooperativeness	0.0112* (0.0047)	0.0315* (0.0131)	-0.0034· (0.0020)	-0.0393* (0.0162)
SVO angle	-0.0027 (0.0032)	-0.0075 (0.0089)	0.0008 (0.0010)	0.0094 (0.0111)
Discount rate	-0.0003 (0.0032)	-0.0010 (0.0090)	0.0001 (0.0010)	0.0012 (0.0113)
Present bias	-0.0050 (0.0032)	-0.0141 (0.0090)	0.0015 (0.0012)	0.0176 (0.0112)
Risk general	0.0061 (0.0041)	0.0171 (0.0113)	-0.0018 (0.0014)	-0.0213 (0.0141)
Risk health	0.0046 (0.0038)	0.0129 (0.0108)	-0.0014 (0.0013)	-0.0161 (0.0134)
Risk financial	-0.0019 (0.0038)	-0.0055 (0.0106)	0.0006 (0.0012)	0.0068 (0.0132)
Trust general	0.0168*** (0.0045)	0.0472*** (0.0120)	-0.0051* (0.0025)	-0.0589*** (0.0147)
Trust family	-0.0056 (0.0039)	-0.0156 (0.0108)	0.0017 (0.0014)	0.0195 (0.0134)
Trust professional	-0.0015 (0.0043)	-0.0043 (0.0121)	0.0005 (0.0013)	0.0054 (0.0151)
Patience	-0.0022· (0.0012)	-0.0061· (0.0033)	0.0007 (0.0004)	0.0075· (0.0041)

Table A4 –Marginal Effect at the mean (MEM) for Not touching face
*Note: · : p < 0.1; *: p < 0.05; **:p < 0.01; ***:p < 0.001 All variables measuring preferences have been normalized prior to regression.*

	1 = “Never”	*2 = “Sometimes”*	*3 = “Often”*	*4 = “Very Often”*
Risk aversion	-0.0046 (0.0050)	-0.0075 (0.0081)	-0.0073 (0.0078)	0.0194 (0.0208)
Cooperativeness	0.0002 (0.0049)	0.0003 (0.0080)	0.0003 (0.0077)	-0.0007 (0.0206)
SVO angle	0.0028 (0.0034)	0.0045 (0.0055)	0.0044 (0.0053)	-0.0116 (0.0142)
Discount rate	0.0034 (0.0034)	0.0055 (0.0055)	0.0053 (0.0053)	-0.0141 (0.0142)
Present bias	-0.0017 (0.0035)	-0.0027 (0.0057)	-0.0026 (0.0055)	0.0069 (0.0147)
Risk general	-0.0015 (0.0043)	-0.0024 (0.0069)	-0.0023 (0.0067)	0.0061 (0.0179)
Risk health	0.0082* (0.0042)	0.0132* (0.0067)	0.0128* (0.0065)	-0.0341* (0.0171)
Risk financial	-0.0014 (0.0040)	-0.0022 (0.0065)	-0.0022 (0.0063)	0.0058 (0.0167)
Trust general	0.0026 (0.0044)	0.0042 (0.0072)	0.0041 (0.0070)	-0.0108 (0.0186)
Trust family	-0.0017 (0.0041)	-0.0028 (0.0066)	-0.0027 (0.0063)	0.0072 (0.0169)
Trust professional	-0.0021 (0.0045)	-0.0034 (0.0074)	-0.0033 (0.0071)	0.0087 (0.0190)
Patience	-0.0016 (0.0012)	-0.0025 (0.0020)	-0.0025 (0.0019)	0.0066 (0.0051)

Table A5 –Marginal Effect at the mean (MEM) for Coughing in sleeves
*Note: · : p < 0.1; *: p < 0.05; **:p < 0.01; ***:p < 0.001 All variables measuring preferences have been normalized prior to regression.*

	1 = "Never"	*2 = "Sometimes"*	*3 = "Often"*	*4 = "Very Often"*
Risk aversion	-0.0007 (0.0005)	-0.0040 (0.0027)	-0.0206 (0.0139)	0.0252 (0.0169)
Cooperativeness	0.0004 (0.0005)	0.0023 (0.0026)	0.0122 (0.0134)	-0.0149 (0.0164)
SVO angle	-0.00001 (0.0003)	-0.0001 (0.0018)	-0.0003 (0.0094)	0.0004 (0.0115)
Discount rate	-0.0002 (0.0003)	-0.0009 (0.0018)	-0.0048 (0.0094)	0.0059 (0.0115)
Present bias	-0.0002 (0.0004)	-0.0012 (0.0020)	-0.0060 (0.0105)	0.0074 (0.0128)
Risk general	0.0010· (0.0006)	0.0061** (0.0025)	0.0320** (0.0120)	-0.0392 (0.0146)
Risk health	0.0003 (0.0004)	0.0016 (0.0021)	0.0081 (0.0107)	-0.0099 (0.0131)
Risk financial	-0.0001 (0.0004)	-0.0006 (0.0021)	-0.0033 (0.0108)	0.0040 (0.0133)
Trust general	-0.0001 (0.0004)	-0.0009 (0.0024)	-0.0044 (0.0123)	0.0054 (0.0150)
Trust family	-0.0004 (0.0004)	-0.0024 (0.0022)	-0.0126 (0.0104)	0.0154 (0.0140)
Trust professional	0.0002 (0.0004)	0.0012 (0.0025)	0.0060 (0.0131)	-0.0073 (0.0160)
Patience	-0.0002 (0.0001)	-0.0010 (0.0007)	-0.0051 (0.0033)	0.0062 (0.0041)

Table A6 –Marginal Effect at the mean (MEM) for Social distancing
*Note: · : p < 0.1; *: p < 0.05; **:p < 0.01; ***:p < 0.001 All variables measuring preferences have been normalized prior to regression.*

	1 = "Never"	*2 = "Sometimes"*	*3 = "Often"*	*4 = "Very Often"*
Risk aversion	-0.0007 (0.0036)	-0.0016 (0.0076)	-0.0020 (0.0096)	0.0043 (0.0208)
Cooperativeness	-0.0018 (0.0035)	-0.0039 (0.0076)	-0.0049 (0.0095)	0.0107 (0.0206)
SVO angle	0.0010 (0.0024)	0.0022 (0.0052)	0.0028 (0.0066)	-0.0061 (0.0142)
Discount rate	-0.0011 (0.0024)	-0.0023 (0.0052)	-0.0029 (0.0065)	0.0062 (0.0142)
Present bias	0.0019 (0.0024)	0.0041 (0.0052)	0.0052 (0.0065)	-0.0112 (0.0140)
Risk general	0.0075* (0.0032)	0.0161* (0.0068)	0.0202* (0.0085)	-0.0438* (0.0182)
Risk health	0.0111*** (0.0032)	0.0238*** (0.0064)	0.0299*** (0.0081)	-0.0648*** (0.0168)
Risk financial	-0.0022 (0.0029)	-0.0047 (0.0062)	-0.0059 (0.0077)	0.0128 (0.0167)
Trust general	0.0024 (0.0032)	0.0051 (0.0069)	0.0064 (0.0087)	-0.0140 (0.0188)
Trust family	-0.0039 (0.0030)	-0.0083 (0.0063)	-0.0104 (0.0080)	0.0225 (0.0172)
Trust professional	0.0012 (0.0033)	0.0026 (0.0071)	0.0033 (0.0089)	-0.0071 (0.0194)
Patience	-0.0008 (0.0009)	-0.0018 (0.0019)	-0.0023 (0.0024)	0.0049 (0.0052)

Table A7 –Marginal Effect at the mean (MEM) for Mask wearing
*Note: · : p < 0.1; *: p < 0.05; **:p < 0.01; ***:p < 0.001 All variables measuring preferences have been normalized prior to regression.*

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Risk aversion	0.0001 (0.0002)	0.0002 (0.0002)	0.0002 (0.0002)	0.0003 (0.0004)	0.0010 (0.0012)	0.0010 (0.0012)	0.0048 (0.0057)	0.0069 (0.0083)	0.0013 (0.0016)	-0.0156 (0.0187)
Cooperativeness	-0.0001 (0.0002)	-0.0002 (0.0002)	-0.0002 (0.0002)	-0.0004 (0.0004)	-0.0011 (0.0012)	-0.0011 (0.0012)	-0.0055 (0.0057)	-0.0080 (0.0082)	-0.0015 (0.0016)	0.0181 (0.0186)
SVO angle	0.0000 (0.0001)	0.0000 (0.0001)	0.0000 (0.0001)	0.0000 (0.0003)	0.0000 (0.0008)	0.0000 (0.0008)	0.0002 (0.0039)	0.0002 (0.0057)	0.0000 (0.0010)	-0.0006 (0.0129)
Discount rate	0.0000 (0.0001)	0.0000 (0.0001)	0.0000 (0.0001)	0.0000 (0.0003)	-0.0001 (0.0008)	-0.0001 (0.0008)	-0.0006 (0.0040)	-0.0009 (0.0057)	-0.0002 (0.0010)	0.0019 (0.0130)
Present bias	0.0000 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0003)	0.0003 (0.0008)	0.0003 (0.0008)	0.0016 (0.0144)	0.0023 (0.0056)	0.0004 (0.0010)	-0.0053 (0.0128)
Risk general	0.0003 (0.0003)	0.0005 (0.0003)	0.0005 (0.0003)	0.0009 · (0.0005)	0.0029 * (0.0012)	0.0029 * (0.0012)	0.0144 ** (0.0052)	0.0207 ** (0.0074)	0.0038 * (0.0018)	-0.0470 ** (0.0165)
Risk health	0.0004 (0.0003)	0.0007 (0.0004)	0.0007 (0.0004)	0.0012 * (0.0006)	0.0038 ** (0.0013)	0.0038 ** (0.0013)	0.0188 *** (0.0049)	0.0271 *** (0.0070)	0.0050 * (0.0019)	-0.0614 *** (0.0153)
Risk financial	0.0001 (0.0001)	0.0001 (0.0002)	0.0001 (0.0002)	0.0002 (0.0003)	0.0007 (0.0010)	0.0007 (0.0010)	0.0033 (0.0046)	0.0047 (0.0067)	0.0009 (0.0012)	-0.0107 (0.0151)
Trust general	0.0004 (0.0003)	0.0006 (0.0004)	0.0006 (0.0004)	0.0010 P (0.0006)	0.0033 * (0.0013)	0.0033 * (0.0013)	0.0163 ** (0.0052)	0.0235 ** (0.0076)	0.0043 * (0.0019)	-0.0533 ** (0.0170)
Trust family	-0.0003 (0.0002)	-0.0004 (0.0003)	-0.0004 (0.0003)	-0.0007 · (0.0004)	-0.0024 * (0.0011)	-0.0024 * (0.0011)	-0.0116 * (0.0049)	-0.0168 * (0.0070)	-0.0031 · (0.0016)	0.0380 * (0.0156)
Trust professional	0.0001 (0.0001)	0.0001 (0.0002)	0.0001 (0.0002)	0.0002 (0.0004)	0.0006 (0.0011)	0.0006 (0.0011)	0.0029 (0.0053)	0.0042 (0.0077)	0.0008 (0.0014)	-0.0095 (0.0174)
Patience	-0.0001 (0.0001)	-0.0002 (0.0001)	-0.0002 (0.0001)	-0.0003 · (0.0002)	-0.0009 * (0.0004)	-0.0009 ** (0.0004)	-0.0046 ** (0.0015)	-0.0066 ** (0.0021)	-0.0012 * (0.0005)	0.0150 ** (0.0047)

Table A8 –Marginal Effect at the mean (MEM) for Respecting Lockdown

*Note: · : p < 0.1; *: p < 0.05; **:p < 0.01; ***:p < 0.001 All variables measuring preferences have been normalized prior to regression.*

Appendix 3: Correlation matrix

	men	<25	>60	self_vul	prox_vul	reg_vul	left	right	rich	poor	day	Risk_av	PGG	SVO	Discount	Present_b	Risk_gen	Risk_he	Risk_fi	Trust_gen	Trust_fam	Trust_pro	patience
men	1 (NA)	-0.045 (0.124)	0.156 (0)	0.038 (0.198)	-0.01 (0.722)	-0.009 (0.75)	0.009 (0.759)	0.113 (0)	0.123 (0)	-0.115 (0)	0.057 (0.053)	0.083 (0.005)	0.085 (0.004)	-0.018 (0.545)	-0.048 (0.105)	-0.044 (0.134)	0.1 (0.001)	0.113 (0)	0.217 (0)	0.062 (0.034)	0.028 (0.349)	0.077 (0.009)	0.011 (0.715)
<25	-0.045 (0.124)	1 (NA)	-0.234 (0)	-0.057 (0.051)	0.03 (0.309)	0.032 (0.283)	0.017 (0.569)	-0.05 (0.088)	-0.099 (0.001)	0.165 (0)	-0.025 (0.404)	-0.083 (0.005)	-0.052 (0.075)	0.029 (0.331)	0.004 (0.905)	-0.024 (0.421)	0.139 (0)	0.134 (0)	0.085 (0.004)	0.011 (0.703)	-0.007 (0.815)	0.034 (0.242)	-0.071 (0.016)
>60	0.156 (0)	-0.234 (0)	1 (NA)	0.153 (0)	0.086 (0.004)	-0.107 (0)	0.058 (0.047)	0.038 (0.195)	-0.019 (0.519)	-0.038 (0.195)	0.002 (0.945)	0.027 (0.351)	0.046 (0.121)	0.035 (0.24)	-0.007 (0.818)	0.004 (0.886)	-0.062 (0.035)	-0.067 (0.022)	-0.009 (0.767)	0.117 (0)	0.045 (0.125)	0.025 (0.396)	0.097 (0.001)
self_vul	0.038 (0.198)	-0.057 (0.051)	0.153 (0)	1 (NA)	0.183 (0)	-0.049 (0.095)	-0.002 (0.938)	0.012 (0.681)	-0.067 (0.022)	-0.004 (0.901)	-0.047 (0.113)	0.013 (0.657)	-0.032 (0.273)	-0.03 (0.308)	-0.018 (0.531)	0.086 (0.003)	-0.105 (0)	-0.095 (0.001)	-0.078 (0.008)	-0.075 (0.01)	-0.055 (0.062)	-0.056 (0.057)	0.019 (0.526)
prox_vuk	-0.01 (0.722)	0.03 (0.309)	0.086 (0.004)	0.183 (0)	1 (NA)	-0.071 (0.015)	0.028 (0.339)	0.007 (0.817)	0.016 (0.578)	-0.056 (0.057)	0.02 (0.489)	-0.014 (0.628)	0.01 (0.722)	0.024 (0.412)	-0.043 (0.141)	0.043 (0.144)	-0.035 (0.233)	-0.033 (0.258)	-0.003 (0.918)	-0.009 (0.764)	-0.043 (0.144)	-0.055 (0.062)	0.054 (0.067)
reg_vul	-0.009 (0.75)	0.032 (0.283)	-0.107 (0)	-0.049 (0.095)	-0.071 (0.015)	1 (NA)	-0.013 (0.662)	0.045 (0.13)	0.08 (0.006)	-0.021 (0.482)	0.028 (0.349)	0.043 (0.142)	0.018 (0.54)	0.02 (0.502)	-0.019 (0.527)	0.033 (0.26)	0.016 (0.591)	0.029 (0.326)	0.052 (0.08)	-0.016 (0.576)	-0.012 (0.69)	0.018 (0.549)	-0.02 (0.495)
left	0.009 (0.759)	0.017 (0.569)	0.058 (0.047)	-0.002 (0.938)	0.028 (0.339)	-0.013 (0.662)	1 (NA)	-0.486 (0)	-0.016 (0.587)	0.02 (0.487)	-0.018 (0.547)	-0.015 (0.609)	0.018 (0.549)	0.104 (0)	-0.05 (0.092)	-0.002 (0.958)	-0.021 (0.474)	0.024 (0.411)	-0.023 (0.425)	0.162 (0)	0.109 (0)	0.179 (0)	0.021 (0.486)
right	0.113 (0)	-0.05 (0.088)	0.038 (0.195)	0.012 (0.681)	0.007 (0.817)	0.045 (0.13)	-0.486 (0)	1 (NA)	0.137 (0)	-0.075 (0.01)	0.038 (0.202)	0.038 (0.197)	0.048 (0.103)	-0.093 (0.002)	-0.002 (0.933)	-0.015 (0.62)	0.065 (0.028)	0.031 (0.29)	0.075 (0.01)	-0.016 (0.579)	-0.02 (0.506)	-0.019 (0.511)	-0.03 (0.311)
rich	0.123 (0)	-0.099 (0.001)	-0.019 (0.519)	-0.067 (0.022)	0.016 (0.578)	0.08 (0.006)	-0.016 (0.587)	0.137 (0)	1 (NA)	-0.361 (0)	0.071 (0.015)	0.113 (0)	0.114 (0)	-0.016 (0.588)	-0.119 (0)	0.011 (0.699)	0.11 (0)	0.068 (0.021)	0.214 (0)	0.146 (0)	0.103 (0)	0.138 (0)	-0.025 (0.389)
poor	-0.115 (0)	0.165 (0)	-0.038 (0.195)	-0.004 (0.901)	-0.056 (0.057)	-0.021 (0.482)	0.02 (0.487)	-0.075 (0.01)	-0.361 (0)	1 (NA)	-0.004 (0.895)	-0.074 (0.012)	-0.05 (0.087)	-0.016 (0.592)	0.066 (0.026)	0.008 (0.795)	-0.042 (0.152)	0.019 (0.516)	-0.11 (0)	-0.119 (0)	-0.106 (0)	-0.189 (0)	0.024 (0.419)
day	0.057 (0.053)	-0.025 (0.404)	0.002 (0.945)	-0.047 (0.113)	0.02 (0.489)	0.028 (0.349)	-0.018 (0.547)	0.038 (0.202)	0.071 (0.015)	-0.004 (0.895)	1 (NA)	0.064 (0.031)	0.091 (0.002)	0.048 (0.101)	-0.016 (0.576)	-0.019 (0.508)	0.023 (0.431)	0.02 (0.505)	0.061 (0.04)	-0.019 (0.529)	-0.006 (0.846)	-0.001 (0.986)	-0.051 (0.08)
Risk_av	0.083 (0.005)	-0.083 (0.005)	0.027 (0.351)	0.013 (0.657)	-0.014 (0.628)	0.043 (0.142)	-0.015 (0.609)	0.038 (0.197)	0.113 (0)	-0.074 (0.012)	0.064 (0.031)	1 (NA)	0.73 (0)	-0.104 (0)	-0.156 (0)	0.004 (0.888)	0.134 (0)	0.046 (0.118)	0.179 (0)	0.096 (0.001)	0.077 (0.008)	0.074 (0.012)	-0.034 (0.247)
PGG	0.085 (0.004)	-0.052 (0.075)	0.046 (0.121)	-0.032 (0.273)	0.01 (0.722)	0.018 (0.54)	0.018 (0.549)	0.048 (0.103)	0.114 (0)	-0.05 (0.087)	0.091 (0.002)	0.73 (0)	1 (NA)	-0.083 (0.005)	-0.166 (0)	-0.011 (0.714)	0.123 (0)	0.033 (0.264)	0.165 (0)	0.091 (0.002)	0.062 (0.035)	0.08 (0.007)	0.003 (0.93)
SVO	-0.018 (0.545)	0.029 (0.331)	0.035 (0.24)	-0.03 (0.308)	0.024 (0.412)	0.02 (0.502)	0.104 (0)	-0.093 (0.002)	-0.016 (0.588)	-0.016 (0.592)	0.048 (0.101)	-0.104 (0)	-0.083 (0.005)	1 (NA)	0.026 (0.377)	-0.032 (0.271)	0.006 (0.849)	0.006 (0.844)	-0.008 (0.797)	0.096 (0.001)	0.088 (0.003)	0.102 (0.001)	0.003 (0.907)
Discount	-0.048 (0.105)	0.004 (0.905)	-0.007 (0.818)	-0.018 (0.531)	-0.043 (0.141)	-0.019 (0.527)	-0.05 (0.092)	-0.002 (0.933)	-0.119 (0)	0.066 (0.026)	-0.016 (0.576)	-0.156 (0)	-0.166 (0)	0.026 (0.377)	1 (NA)	0.046 (0.12)	-0.053 (0.074)	-0.034 (0.248)	-0.077 (0.009)	-0.064 (0.03)	-0.086 (0.003)	-0.081 (0.006)	-0.048 (0.105)
Present_b	-0.044 (0.134)	-0.024 (0.421)	0.004 (0.886)	0.086 (0.003)	0.043 (0.144)	0.033 (0.26)	-0.002 (0.958)	-0.015 (0.62)	0.011 (0.699)	0.008 (0.795)	-0.019 (0.508)	0.004 (0.888)	-0.011 (0.714)	-0.032 (0.271)	0.046 (0.12)	1 (NA)	-0.003 (0.931)	0.014 (0.635)	0.013 (0.659)	-0.016 (0.577)	0.008 (0.789)	-0.033 (0.269)	0.006 (0.84)
Risk_gen	0.1 (0.001)	0.139 (0)	-0.062 (0.035)	-0.105 (0)	-0.035 (0.233)	0.016 (0.591)	-0.021 (0.474)	0.065 (0.028)	0.11 (0)	-0.042 (0.152)	0.023 (0.431)	0.134 (0)	0.123 (0)	0.006 (0.849)	-0.053 (0.074)	-0.003 (0.931)	1 (NA)	0.55 (0)	0.48 (0)	0.237 (0)	0.134 (0)	0.208 (0)	-0.09 (0.002)
Risk_he	0.113 (0)	0.134 (0)	-0.067 (0.022)	-0.095 (0.001)	-0.033 (0.258)	0.029 (0.326)	0.024 (0.411)	0.031 (0.29)	0.068 (0.021)	0.019 (0.516)	0.02 (0.505)	0.046 (0.118)	0.033 (0.264)	0.006 (0.844)	-0.034 (0.248)	0.014 (0.635)	0.55 (0)	1 (NA)	0.399 (0)	0.199 (0)	0.049 (0.1)	0.172 (0)	-0.093 (0.002)
Risk_fin	0.217 (0)	0.085 (0.004)	-0.009 (0.767)	-0.078 (0.008)	-0.003 (0.918)	0.052 (0.08)	-0.023 (0.425)	0.075 (0.01)	0.214 (0)	-0.11 (0)	0.061 (0.04)	0.179 (0)	0.165 (0)	-0.008 (0.797)	-0.077 (0.009)	0.013 (0.659)	0.48 (0)	0.399 (0)	1 (NA)	0.193 (0)	0.091 (0.002)	0.196 (0)	-0.061 (0.037)
Trust_gen	0.062 (0.034)	0.011 (0.703)	0.117 (0)	-0.075 (0.01)	-0.009 (0.764)	-0.016 (0.576)	0.162 (0)	-0.016 (0.579)	0.146 (0)	-0.119 (0)	-0.019 (0.529)	0.096 (0.001)	0.091 (0.002)	0.096 (0.001)	-0.064 (0.03)	-0.016 (0.577)	0.237 (0)	0.199 (0)	0.193 (0)	1 (NA)	0.481 (0)	0.607 (0)	0.016 (0.58)
Trust_fam	0.028 (0.349)	-0.007 (0.815)	0.045 (0.125)	-0.055 (0.062)	-0.043 (0.144)	-0.012 (0.69)	0.109 (0)	-0.02 (0.506)	0.103 (0)	-0.106 (0)	-0.006 (0.846)	0.077 (0.008)	0.062 (0.035)	0.088 (0.003)	-0.086 (0.003)	0.008 (0.789)	0.134 (0)	0.049 (0.1)	0.091 (0.002)	0.481 (0)	1 (NA)	0.524 (0)	-0.004 (0.9)
Trust_pro	0.077 (0.009)	0.034 (0.242)	0.025 (0.396)	-0.056 (0.057)	-0.055 (0.062)	0.018 (0.549)	0.179 (0)	-0.019 (0.511)	0.138 (0)	-0.189 (0)	-0.001 (0.986)	0.074 (0.012)	0.08 (0.007)	0.102 (0.001)	-0.081 (0.006)	-0.033 (0.269)	0.208 (0)	0.172 (0)	0.196 (0)	0.607 (0)	0.524 (0)	1 (NA)	-0.008 (0.782)
patience	0.011 (0.715)	-0.071 (0.016)	0.097 (0.001)	0.019 (0.526)	0.054 (0.067)	-0.02 (0.495)	0.021 (0.486)	-0.03 (0.311)	-0.025 (0.389)	0.024 (0.419)	-0.051 (0.08)	-0.034 (0.247)	0.003 (0.93)	0.003 (0.907)	-0.048 (0.105)	0.006 (0.84)	-0.09 (0.002)	-0.093 (0.002)	-0.061 (0.037)	0.016 (0.58)	-0.004 (0.9)	-0.008 (0.782)	1 (NA)

Table A9: Correlation between explanatory variables. (first line indicates Pearson correlation coefficient, second line in parenthesis indicates the p-value of the Pearson test).

Appendix 4 : Variable selected for each predictive model

		Washing hands					Avoiding touching face					Coughing into sleeves					Social distancing					Mask wearing					Lockdown compliance				
		linreg	sir	plsr	ridge	olm	linreg	sir	plsr	ridge	olm	linreg	sir	plsr	ridge	olm	linreg	sir	plsr	ridge	olm	linreg	sir	plsr	ridge	olm	linreg	sir	plsr	ridge	olm
Elicited preferences	Risk aversion	10%	6%	11%	6%	19%	48%	31%	47%	**51%**	**73%**	5%	10%	5%	3%	19%	18%	13%	18%	9%	45%	14%	11%	13%	8%	31%	8%	10%	8%	8%	24%
	Cooperativeness	**52%**	29%	**51%**	36%	**51%**	**68%**	50%	**68%**	**68%**	**90%**	11%	12%	11%	7%	34%	8%	8%	8%	4%	23%	10%	8%	9%	5%	24%	13%	13%	13%	14%	40%
	SVO angle	5%	6%	5%	4%	6%	5%	9%	5%	6%	11%	36%	23%	36%	30%	34%	4%	5%	4%	3%	6%	38%	23%	37%	34%	28%	4%	6%	4%	4%	4%
	Discount rate	38%	22%	37%	32%	36%	8%	8%	8%	6%	9%	5%	12%	4%	4%	5%	6%	8%	7%	7%	12%	11%	6%	12%	11%	11%	9%	7%	12%	8%	7%
	Present bias	3%	1%	3%	2%	10%	**59%**	38%	**58%**	**52%**	**55%**	17%	16%	18%	9%	11%	6%	2%	6%	8%	7%	18%	9%	19%	17%	16%	5%	3%	6%	6%	9%
Stated preferences	Risk_general	2%	4%	2%	3%	10%	10%	9%	12%	5%	31%	**57%**	36%	**60%**	**51%**	**68%**	**82%**	**57%**	**82%**	**62%**	**94%**	**89%**	**76%**	**90%**	**75%**	**94%**	**98%**	**78%**	**97%**	**88%**	**97%**
	Risk_health	**98%**	**93%**	**98%**	**95%**	**99%**	30%	20%	31%	16%	34%	**82%**	**66%**	**82%**	**75%**	**89%**	22%	21%	22%	7%	42%	**100%**	**85%**	**99%**	**92%**	**99%**	**100%**	**100%**	**100%**	**99%**	**100%**
	Risk_financial	3%	2%	3%	1%	6%	8%	10%	6%	6%	15%	7%	12%	8%	5%	15%	11%	7%	12%	4%	12%	17%	12%	16%	20%	26%	15%	12%	14%	8%	30%
	Trust_general	33%	20%	34%	21%	**58%**	**100%**	**98%**	**100%**	**100%**	**100%**	**68%**	42%	**67%**	**55%**	**72%**	5%	3%	5%	3%	15%	4%	3%	4%	5%	12%	**96%**	**83%**	**95%**	**93%**	**98%**
	Trust_family	**53%**	29%	**55%**	47%	**79%**	47%	29%	49%	47%	**69%**	27%	25%	25%	20%	35%	44%	33%	43%	29%	**72%**	39%	26%	38%	37%	**79%**	**98%**	**90%**	**98%**	**96%**	**98%**
	Trust_pro	12%	5%	11%	9%	30%	7%	9%	5%	4%	14%	12%	14%	11%	11%	21%	4%	3%	3%	5%	13%	12%	7%	14%	11%	14%	9%	5%	8%	7%	18%
	Patience	7%	7%	6%	8%	11%	**69%**	42%	**69%**	**60%**	**70%**	4%	11%	4%	6%	7%	43%	32%	44%	32%	**58%**	35%	15%	35%	34%	37%	**87%**	**52%**	**84%**	**84%**	**95%**
Individual caracteristics and controls	Male	**100%**	**100%**	**100%**	**100%**	**100%**	**99%**	**96%**	**100%**	**100%**	**99%**	**100%**	**96%**	**100%**	**100%**	**100%**	17%	11%	20%	14%	28%	**99%**	**94%**	**100%**	**99%**	**99%**	**98%**	**82%**	**99%**	**97%**	**100%**
	< 25	**67%**	50%	**68%**	44%	**68%**	**72%**	46%	**73%**	50%	**67%**	15%	16%	15%	5%	39%	**100%**	**98%**	**100%**	**100%**	**100%**	13%	9%	12%	1%	9%	**63%**	**52%**	**62%**	35%	**72%**
	> 60	**98%**	**91%**	**97%**	**94%**	**99%**	**53%**	35%	**52%**	46%	**64%**	**100%**	**99%**	**100%**	**100%**	**100%**	**100%**	**94%**	**100%**	**99%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	16%	11%	13%	9%	16%
	high income	10%	9%	12%	5%	11%	**98%**	**91%**	**99%**	**94%**	**99%**	13%	17%	11%	4%	24%	6%	8%	4%	2%	23%	26%	11%	27%	17%	39%	**67%**	40%	**66%**	**54%**	**75%**
	low income	2%	3%	3%	2%	6%	**91%**	**72%**	**92%**	**76%**	**91%**	31%	21%	31%	13%	**55%**	24%	9%	25%	11%	13%	20%	4%	20%	8%	8%	5%	6%	5%	2%	10%
	left opinions	7%	9%	6%	3%	22%	3%	7%	4%	2%	9%	4%	12%	4%	4%	7%	9%	5%	9%	5%	13%	28%	27%	25%	17%	31%	6%	6%	5%	2%	12%
	right opinions	11%	10%	11%	7%	15%	5%	9%	6%	3%	11%	3%	9%	4%	1%	12%	8%	4%	10%	6%	10%	6%	5%	7%	2%	8%	4%	7%	5%	1%	9%
	Vulnerable person	3%	4%	2%	1%	9%	18%	13%	19%	10%	17%	37%	26%	37%	20%	45%	**92%**	**67%**	**94%**	**84%**	**94%**	**95%**	**65%**	**95%**	**92%**	**98%**	46%	26%	46%	32%	**64%**
	Liv. with. vuln. pers.	18%	13%	19%	12%	24%	4%	10%	3%	1%	5%	13%	13%	13%	5%	6%	12%	9%	13%	6%	16%	9%	14%	11%	5%	30%	**56%**	30%	**53%**	36%	39%
	Highly infected area	10%	8%	11%	5%	18%	4%	8%	4%	1%	8%	14%	15%	13%	5%	17%	7%	7%	6%	3%	17%	9%	5%	8%	4%	4%	**97%**	**78%**	**97%**	**93%**	**95%**
	Survey date	13%	9%	15%	15%	15%	44%	27%	45%	44%	39%	35%	22%	34%	35%	30%	27%	16%	28%	26%	47%	**73%**	46%	**71%**	**74%**	**62%**	46%	35%	45%	**51%**	32%

Table A10: Variable selection rate for each prophylactic measure and each regression model.