



## Big Data is Decision Science: the Case of Covid-19 Vaccination

### Authors

#### **Jacques Bughin**

Solvay Brussels School of Economics and Management, Université Libre de Bruxelles,  
iCite, Brussels, MachaoneAdvisory, Brussels, FortinoCapital, Brussels, Antler, Amsterdam and  
Portulans Institute, Washington - email : : [bughinjacquesrenejean@gmail.com](mailto:bughinjacquesrenejean@gmail.com)

#### **Michele Cincera**

Solvay Brussels School of Economics and Management, Université libre de Bruxelles,  
iCite and ECARES - email: [mcincera@ulb.ac.be](mailto:mcincera@ulb.ac.be)

#### **Dorota Reykowska**

Neurohm, Poland – email: [Dorota.Reykowska@neurohm.com](mailto:Dorota.Reykowska@neurohm.com)

#### **Rafal Ohme**

Neurohm & WSB, University, Torun, Poland – email : [Rafal.Ohme@umk.pl](mailto:Rafal.Ohme@umk.pl)

**iCite Working Paper 2021-047**

# Big data is decision science: the case of covid-19 vaccination

Jacques Bughin

*Université libre de Bruxelles, Solvay Brussels School of Economics and Management, iCite and ECARES, Belgium*

*Machaon Advisory, Brussels, FortinoCapital, Brussels, Antler, Amsterdam and Portulans Institute, Washington*

Michele Cincera

*Université libre de Bruxelles, Solvay Brussels School of Economics and Management, iCite and ECARES, Belgium*

Dorota Reykowska

*NEUROHM, Poland*

Rafał Ohme

*WSB University, Poland*

## ABSTRACT

*Data science has been proven to be an important asset to support better decision-making in a variety of settings, whether it is for a scientist to better predict climate change, for a company to better predict sales, or for a government to anticipate voting preferences. In this research, we leverage Random Forest (RF) as one of the most effective machine learning techniques using big data to predict vaccine intent in five European countries. The findings support the idea that outside of vaccine features, building adequate perception of the risk of contamination, as well as securing institutional and peer trust are key nudges to convert skeptics to get vaccinated against the covid-19. What machine learning techniques further add beyond traditional regression techniques, is some extra granularity in factors affecting vaccine preferences (twice more factors than logistic regression). Other factors that emerge as predictors of vaccine intent are compliance appetite with non-pharmaceutical protective measures, as well as perception of the crisis duration.*

Keywords: Attitudes, Big data, Covid-19, iCode™, Machine learning techniques, Random Forest, Response time, Vaccination,

## INTRODUCTION

Big data is the handling of vast amounts of data through a flexible, mostly cloud-based IT architecture. Big data is here to stay for many reasons. The first is that digitization is data extension, with new digital content creation growing at a rate between 40 and 60% a year. This rate means that Google, which was indexing a million pages for a few million searches in 1998, now indexing more than a trillion pages ten years later, for 1.2 trillion searches a year. Second, big data can fuel new powerful machine learning techniques to uncover otherwise hidden relationships between data, that can support new powerful insights. As an example, Netflix shifted its content recommendation engine, based on customer rating, to a machine-learning algorithm fed by a large set of big data ten years ago. The tool offered a revolution to support personalized recommendations, in such a way, that now, four out of five Netflix movies and TV series viewed by Netflix subscribers originate from machine-based suggestions (Amatriain, 2013).

The seminal work by Brynjolfsson, et al. (2011) reveal that companies leveraging big data for more fine-tuned decision making, could increase their revenue productivity by more than 5%. A few years

later and using a more global sample, Bughin (2016) had reached essentially the same conclusion- that is big data is usually responsible for an uplift of more than 5% in labor productivity for firms globally. Besides being used for business, big data is also important when it comes to helping major social issues, in particular when it comes to complex issues such as predictions of climate change (Bauer et al., 2021), vote intent (Mavragani and Tsagarakis, 2019), or traffic congestion (Teseng and colleagues 2018).

One another current case in point is the vaccination issue linked to the covid-19 pandemic. While the covid-19 pandemic continues unabated across the world, affecting hundreds of million people worldwide, the only serious way to eradicate the pandemic is mass- vaccination. However, large vaccination acceptance is not warranted. Social movements against vaccination have increased during the covid-19 pandemic, fueled by a series of conspiracy theories (Nguyen and Catalan, 2020). There are also possibly multiple, and interacting elements that affect vaccine intent, which make predictions of vaccine used complex—A recent poll performed by the high profile organization Pew Research in the US also emphasizes that the shape of vaccine intent is driven by « rather complex and interrelated factors » that might require more deeper data analytics to sort out a clear ranking of actions to support mass-vaccination<sup>1</sup>.

We thus leverage machine learning to profile vaccination intent for covid-19, concentrating outside of known product factors (such as the effectiveness and safeness of the vaccine) to determine additional factors that could be boosted or alleviated to enhance fast vaccination against the Covid-19. Specifically, after aligning with recent theories of vaccine determinants to build an extensive database of behaviors and vaccine intent across 5 Continental European countries—Germany, France, Spain, Italy and Sweden, we use Random Forest (RF) techniques to provide a granular view of vaccine intent. RF is one of many machine learning techniques, but is found to generate rather accurate predictions regarding protective uptake (Bughin and Cincera, 2021)

Reaching for advanced analysis techniques has been complemented by a dedicated approach to data collection. Predicating behavior in a sensitive context such as vaccination intent in the face of the covid-19 pandemic required the usage of tools that would help decrease the natural tendency to distort declarations.

Understanding the real motives of behavior has always been a challenge for researchers. Numerous studies have shown that there is a weak correlation between declarations and behavior (Krauss, 1995) making it hard to predict behavior based solely on explicit answers. Especially when testing attitudes around sensitive topics, full of emotional load. Political correctness, post rationalizations or auto presentation needs are all important aspects that frequently influence and distort explicit, declarative answers. Additionally in the last 50 years researchers have consistently shown that most of our cognitive processes take place outside of conscious awareness and control and still they influence our perceptions, judgments and actions (Zajonc, 1968, 1980; Uleman & Bargh 1989; Bornstein & Pittman 1992; Greenwald 1992; Murphy & Zajonc 1993; Bargh 1997; Ohme, 2001; Nosek, Hawkins, & Frazier, 2011). For this reason in the project dedicated to understanding covid-19 vaccination intent we decided to use Response Time measurement and follow the framework of Fazio's attitude accessibility theory. Based on Fazio there are 2 components of attitudes - explicit opinion and implicit accessibility of attitude (Fazio, 2001). Attitudes that are stronger, are more accessible and therefore expressed with a shorter response time (Fazio & Williams, 1986; Fazio et al., 1989). Fazio, Powell, and Williams (1989) posited that RT might be an indicator for an attitude strength - the more quickly an attitude is expressed, the greater its strength. Stronger and more accessible attitudes have a stronger influence on behavior.

The findings are threefold. The first is the confirmation that machine learning techniques confirm a set of factors affecting the mindset of citizens with regards to vaccine intent. Those factors are in line with theories and early findings.

Second, by choosing 5 countries which were in different stages of impact, and of policy reactions to the covid-19 pandemic, we are able to sort out common and country-specific factors, according to pandemic stage and its perception by citizens. For instance, by April 2020, Sweden and Germany had suffered much lower contagion and had adopted more relaxed social distancing policies than Spain,

---

<sup>1</sup> [Intent to Get a COVID-19 Vaccine Rises to 60% as Confidence in Research and Development Process Increases | Pew Research Center](#)

Italy and France; hence, worries about health and home isolation that typically sort out people regarding their vaccination intent, is not emerging as such discriminant power in those countries. We also notice some differences by countries, such as media perception (lifting vaccine intent, if media exaggerate the situation).

Finally, what big data techniques bring on top of other techniques (such as multilogit regressions), is the granularity of insights; the techniques are able to point out to a set of factors that are more country- and vaccine intent type- specific, which otherwise would have remained “hidden” in the data. As an example, logit regression would show that those citizens satisfied by how their government handles the crisis have a higher propensity to get vaccinated; what big data techniques add, is that high satisfaction brings higher vaccination intent in all countries, except Italy, and that dissatisfaction correlates more with hesitancy in strict social distancing countries, than in Sweden and Germany, where dissatisfaction is only correlated with refusal to be vaccinated. This level of granularity is thus important to target the right segments of citizens.

The research reads as follows- first, we discuss the research background; then we present results of the Random Forest techniques. The last section concludes.

## RESEARCH BACKGROUND

This research derives from a project at assessing citizen’s attitudes, and behaviors linked with the SARS-CoV-2 pandemic, and how those aspects influence protective behavior. Bughin et al., (2020) concentrate on non-pharmaceutical protection interventions (NPIs). This research focuses on *vaccination*. The data from this survey were collected in April 2020, during the first wave of the covid-19, in France, Germany, Italy, Spain and Sweden. Data were assembled online and rely on a country representative sample for age and gender. Each country sample obeys the restriction of > 1,000 answers per country. For all questions, respondents had the choice to answer “yes, hesitant, no”. Note as well that intent rates have been adjusted by response time (RT), as collected through the iCode Smart test (Ohme et al., 2020). The procedure amounts to re-center the probability of acceptance and refusal towards more hesitancy, as too quick or too long response time may indeed reduce the credibility of answers given by the respondent. About 7 points out of the 22% percent arise out of the adjusted procedure, but this should not be considered as an artifact, rather an appropriate adjustment to ensure that intent is a good predictor of actual behavior (Ohme et al., 2020). iCode™ smart test is a web-based and device agnostic technology created by NEUROHM. Apart from declarations, it captures true attitudes which are free from conformity, social and cognitive biases or wishful thinking. By measuring response time (RT) iCode™ estimates how much people hesitate when they express their opinion in a survey. Higher confidence, expressed by shorter response time indicates a well-established, internalized attitude that is more likely to drive behavior (Fazio, 2001).

For respondents the task is very simple. They are asked to evaluate if they agree with the statements presented on the screen. The answers are given on a 3 point scale (yes, hard to tell, no) and response time (RT) is measured for each answer, making it possible to collect at the same time both explicit (declarative) answers as well as implicit (indirect) attitudes.

To ensure high quality of data and eliminate test biases a calibration phase and control screen are added. Calibration precedes the test phase and consists of 3 steps:

- Familiarization with the scale
- Familiarization with the purpose of the task
- Increasing the focus on the task

Additionally a control screen is introduced to eliminate the effect of the position of the mouse on the screen. It is presented before each statement, forcing a standardized position of the mouse (the distance to the yes and no answers is always the same).

## Vaccination Intent

Table 1 displays vaccination intent by countries. The share of vaccination is 65% across all countries, but with a large spread, e.g. acceptance is only 56% among Swedish citizens, and up to 74% in Spain. Furthermore, for each country, we notice that the portion of “hesitants” is larger than the share of “refusers” and more than twice in Spain for instance. The refusers segment is the largest in Germany. Vaccination intent is in line with other research, that indeed shows on average low acceptance rate in European countries and the US (see Sallam, M. (2021), Ruiz & Bell (2021) and Attwell et al. (2021)). What is even more important is that intent may be too low to ensure herd immunity with certainty. E.g. in the context of a full homogenous transmission, and a reproduction rate for covid-19, at a mode of  $R_0$  at 3, herd immunity is achieved at  $1-1/3=66\%$ . Eissler (2021), furthermore shows that for  $R_0$  above 2.5, the probability of herd immunity to be achieved, even if infection transmission is stochastic is less than 10%, thus highlighting the risk that vaccination intent today does not warrant herd immunity for sure, and thus prevent the removal of NPI as a way to limit the contagion. In such a context, we see thus as well the importance to clearly spot the drivers that would convert the hesitants to vaccination.

Table 1. Number of Covid-19 European respondents, April 2020

Countries	Willingness to be vaccinated		
	Yes	Hesitant	Refusers
France	0.56	0.28	0.16
Germany	0.63	0.20	0.17
Italy	0.71	0.19	0.10
Spain	0.74	0.18	0.08
Sweden	0.56	0.28	0.16

Note: All variables are corrected by Response Time (RTC)

## Conceptual model of vaccination intent

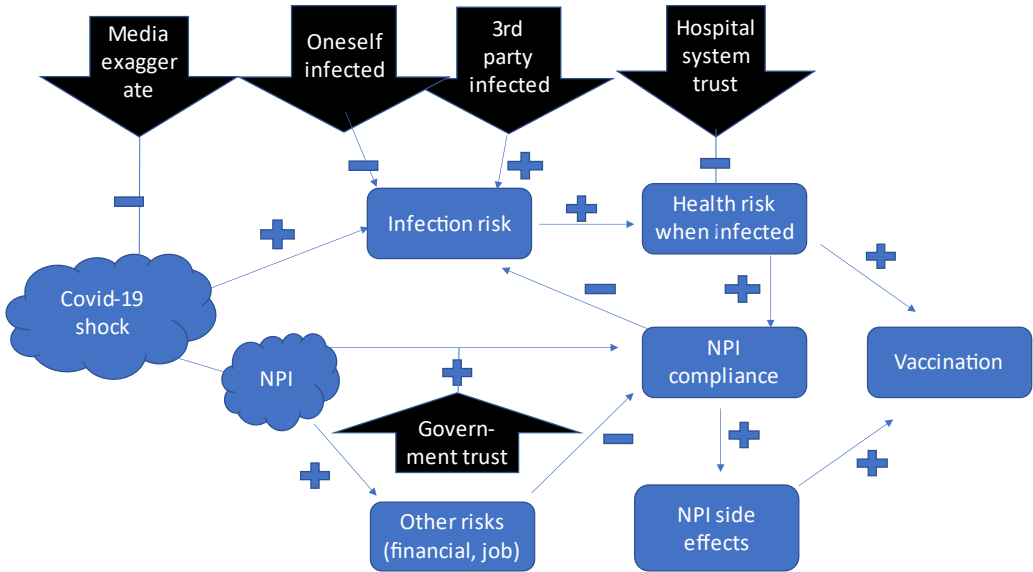
We piggy back on multiple research and theories to inform about the key variables to possibly play a role on vaccine intent.

The conceptual model in mind is presented in Figure 1. The conceptual model puts asides the importance of key product attributes for acceptance. As our survey was designed before the various vaccines were discovered, we have based our research on the assumption that the vaccine is highly effective, with limited side effects. While not explicitly mentioned, the conceptual model is also mediated by disposition of individuals, which relates to certain types of socio-demographics, e.g. research informs of possible difference in vaccination behavior among gender, even if largely contextual. Older adults and individuals with higher education and income also show higher vaccination intention than the average, -and for a large set of vaccine uptake (Larson et al., 2014). In the case of covid-19, we state that age is especially important as the risk of lethal infection increases significantly after 60 years old. As this age matches the modal time when European citizens tend to retire from the workforce, age also impacts other financial and job sustainability risks in our conceptual model.

The conceptual model dictates that the speed of the covid-19 pandemic shock hitting most of Europe by March 2020 may have led to a worry by the population of a) being infected- and, in such a case b) of suffering health issues related to catching the virus, - a fortiori for those mistrusting the ability of the hospital care to manage the disease. The risk perception is however not uniform among citizens and depends on their indirect experience provided by the media or by the evidence of the pandemic and its health effects, among close social ties. Those who have been already infected and have

recovered may adapt their risk perception, but also will have lower incentive to get vaccinated, as they have antibodies protection for a while. Also, the covid-19 shock was also accompanied by a series of more or less stringent NPI, from social distancing to quarantines. Those measures will be a fortiori more respected by those perceiving the risk on their health, as well by those obeying to social norms and /or trusting the government actions. Stringent NPI also leads to two types of negative externalities. The first, direct effect is the large economic shutdown affecting those who may be prevented to work; the second more long-term effect has been the bad-being impact of isolation, and/or too much proximity leading to family tension. The first externality would likely lead to lower NPI compliance; the second effect is a negative burden out of NPI compliance, leading to a shift towards more vaccination.

Figure 1. Conceptual model of vaccination



The prediction towards both NPI behavior, and vaccine intent, as a result of a pandemic shock risk perception, is based on the health belief model pioneered by Becker (1974) and discussed further in Brewer et al. (2007). Here, we also split risk in terms of getting infected, and the morbidity risk associated with being infected. We explicitly consider importance of governance and communications linked to major moral hazards, as this is often seen as a major booster of risk perception and actions, see Wachinger et al. (2013). Here we consider perceptions of citizens shaped by media communications, but also by trust in policy actions (especially NPIs) and trust in health systems quality. We finally consider the importance of social norms—e.g. health risk perception by individuals can be shaped by close family, friends. People also often try to influence others’ to follow their behavior in the context of major disruptive events, like floating, earthquakes or pandemic (see Yang, et al. 2018). Finally, protective interventions can be pharmaceutical or not. As NPIs are known to have large negative externalities, some people would want to trade them for faster vaccination. Such tradeoff has been shown to prevail for Germany in a set of specific conjoint analyses (Bughin et al. 2021).

Table 2 provides the list of variables collected that match the model design. We have close to 50 behavioral, emotions and attitudinal questions, on top of about 10 questions linked to socio-demographics, for more than 5000 respondents. This leads to a database of more than 300,000 data points, or a « big data » set to play with. As also seen in Table 2, the considered NPIs measures include “item disinfection” (n4 and n5), or “hands cleaning” (n6), home stay and physical distancing

(n2 and n3). The last two are typically the most important ones when it comes to assessing the link with health risk perception (Cho, 2020).

Regarding entities trust, we used multiple variables, e.g. for media, we look at perception of media communication exaggeration, as well as of its provision of reliable information; likewise we consider perception of government actions and information transparency. Referring to healthcare, we focus not only on how the health system manages the crisis, but also whether people perceive the health system as providing appropriate help. For instance, the perception can be that the health system has enough bed capacity for complex cases, but that the help is ineffective in intensive care, as the death rate significantly increase for people being hospitalized and getting to intensive care units.

Health worry is the largest worry expressed among citizens (62% of the total sample), but other worries as a result of shutdown were large too, e.g. job preservation was perceives as a challenge for 49% of the respondents. Isolation worries are also important, especially in countries with strict social distancing rules, during the first wave of the pandemic, like Spain and Italy. The general sense was also that people were complying with NPI, even if only 55% of citizens believed that media has been providing reliable information about the pandemic. 62% believe that the health system is adequately handling this crisis, while only 55% were satisfied with how the government is handling the pandemic crisis. Last but not least, more than 7 out of 10 people responded that they were complying with physical distancing, and home quarantine. This level is high, but far from complete and is consistent with other works, e.g. Zickfeld et al. (2020).

*Table 2. Variables measured in all 5 countries*

Variable	<b>BEHAVIOR</b>
RTC.n1	I actively encourage others to follow the restrictions and guidelines
RTC.n2	I comply with the recommendations for physical distancing
RTC.n3	I comply with the restrictions to stay home
RTC.n4	I disinfect groceries before putting them away
RTC.n5	I disinfect mail and deliveries before opening them
RTC.n6	I wash hands for 20 seconds when necessary
RTC.n7	I would like to help people who are more vulnerable to COVID-19
RTC.n8	Since COVID-19 I eat healthier
RTC.n9	Since COVID-19 I eat unhealthier
RTC.n10	Since COVID-19 I exercise less
RTC.n11	Since COVID-19 I exercise at home more
	<b>EMOTIONS</b>
RTC.n13	I'm worried about my financial situation
RTC.n14	I'm worried about my job situation
RTC.n15	I'm worried that our country will run out of money
RTC.n16	I'm worried that there will not be enough basic necessities in the stores
RTC.n17	I am worried about my own health
RTC.n18	I am worried about the health of my children
RTC.n19	I am worried about the health of my older family members
RTC.n20	I am worried about the health of people in my country
RTC.n21	I worry that there will be an increase in break-ins and thefts
RTC.n22	I'm worried about my children's education
RTC.n23	I am anxious about not being able to meet with friends
RTC.n24	I am worried about not being able to meet with my family
RTC.n25	I worry how living in isolation will affect me
RTC.n26	Living in isolation negatively impacts my wellbeing
	<b>OPINIONS</b>
RTC.n27	The COVID-19 outbreak will make society more unequal
RTC.n28	Being together all the time increases family tensions
RTC.n29	COVID-19 increases domestic violence
RTC.n30	COVID-19 will increase divorce rates
RTC.n31	COVID-19 will bring countries closer
RTC.n32	I am grateful to our essential workers

RTC.n33	I am grateful to our healthcare professionals
RTC.n34	My chance of getting COVID-19 is high
RTC.n35	Slowing the spread of COVID-19 is more important than the economy
RTC.n36	Coronavirus is dangerous for my health
RTC.n37	Media exaggerate the situation with COVID-19
RTC.n38	Media provide reliable information about the pandemic
RTC.n39	[The President] is doing a good job dealing with COVID-19
RTC.n40	I am satisfied with how my government is handling this crisis
RTC.n41	The government is doing a good job dealing with COVID-19
RTC.n42	I am satisfied with how our healthcare system is handling this crisis
RTC.n43	In the case of coronavirus infection, I will get appropriate medical help
RTC.n44	The government discloses real numbers of coronavirus infections and deaths
RTC.n45	COVID-19 reveals the best in people
RTC.n46	COVID-19 reveals the worse in people
RTC.n47	I believe we will beat COVID-19 soon
RTC.n48	People will stop following the restrictions soon

In general, we found that differences between countries are both cultural and contextual. In a country with more social accountability like Sweden, “helping people who are vulnerable to covid-19” is an important predictor of vaccine preferences, while it is not in other countries, except Spain. In Germany, a country known, for respecting rules, compliance to NPI is more closely linked to vaccine uptake than other countries. Family tensions emerging from NPIs are naturally higher in Italy and Spain than in other countries as the latter had put less stringent distancing rules than the former.

## TO BE OR NOT TO BE VACCINATED- MACHINE LEARNING INSIGHTS

This section formally tests the conceptual model.

### Traditional models

As a first step, we have run a simple univariate correlation between vaccine intent and all other variables. The exercise (not reproduced here for sake of space) shows that among all variables of Table 2, only a restricted amount of variables (about 6) exhibit significant correlation with vaccination propensity. In effect, vaccination appears to be *negatively* linked with the statement that “media exaggerate the situation with COVID-19”, while it positively correlates with various indicators of health worries, (worry about one’s, as well as about own children’s’ and older family members’ health). Age, especially, 64+ older people, correlates with higher intent to get vaccinated. While the sign of the correlation fits with the literature, the univariate correlation analysis does not suggest a) any role for NPI and b) for government, healthcare and peer actions to shape vaccine preferences. Neither contextual factor such as income, nor other risk externalities beyond health, such as isolation, financial risk etc. appear to influence vaccine intent.

As a second step and further to univariate correlation analysis, we also have performed a multilogit regression on vaccine intent preferences. The high-level results are synthesized in Tables 3 and 4 for all countries pooled together. Table 3 lays out the variables that are statistically significant at 5%, while, for ease of lecture, Table 4 aggregates various indicators to fit with the conceptual model, and computes marginal impact on vaccination<sup>2</sup>.

Table 3. Logit estimates, vaccination intent, April 2020, pooled for all 5 countries

VARIABLES	YES	HST	NO
26-35		0.433**	
36-49	-0.583***	0.491***	
50-64	-0.662***	0.490***	0.429*

<sup>2</sup> Note that only statistically significant regression coefficients are displayed and added as marginal impact to vaccine effects. We nevertheless have taken aside significant country fixed effects as well as demographic effects



>64	-0.519***	0.502***	
>2000 euros/month		-0.247**	
Exposed/Not to Covid		-0.345**	
Not exposed			0.420**
I would like to help people who are more vulnerable to COVID-19	0.734***	-0.660***	
Since COVID-19 I eat unhealthier			-0.448*
I'm worried about my job situation		-0.387*	
I am worried about my own health	0.316**		-0.728***
Being together all the time increases family tensions	0.409**	-0.392*	-
COVID-19 increases domestic violence		-0.276	0.455*
COVID-19 will increase divorce rates			-0.644**
I am grateful to our essential workers	0.652***		-0.794***
I am grateful to our healthcare professionals			-0.726**
My chance of getting COVID-19 is high	1.109***	-0.748***	-0.681*
Coronavirus is dangerous for my health	0.839***	-0.290*	-0.809***
Media exaggerate the situation with COVID-19	-0.830***		1.122***
Media provide reliable information about the pandemic	0.413**		-0.781***
In the case of coronavirus infection, I will get appropriate medical help	1.187***	-0.707***	-0.976***
The governments disclose the real numbers about Covid	0.494**		-1.091***

Notes:  $Prob > F = 0,000$ ,  $pseudo R^2 = 0.23$

The logistic regression is able to capture much more effect than simple one-way correlation. In fact, 17 factors out of our list come out to be statistically significant. Further, the statistically significant variables affecting vaccine intent tend to match the conceptual model, and the derived marginal effects on vaccine preferences (ref. Table 4) are about in line with other literature estimates:

1. As said, informational factors should play a role, as at the start of the covid-19 pandemic, not so many people were contaminated, and the virus was new with large uncertainty as to its likely effects, so that people would need to base their risk perception on external clues. Two effects seem to play, but in opposite directions. Not witnessing close ties being infected plays a negative role on propensity to get vaccinated (13% lower intent of getting vaccinated), while the fact that media can provide reliable information increases the vaccine intent by the same margin.
2. Health matters have the largest direct positive effect on vaccine intent- by combining both risks of being contaminated and health issue if infected, the probability of getting vaccinated increases by 22%.
3. NPI measures do not appear to boost vaccine intent, in contrary to NPI indirect side effects. When the effects are adverse, people seek vaccination in order to get out of their situation; but the effects are asymmetric, e.g. domestic tensions affects will to vaccinate, low risk of divorce reduces will to vaccinate.
4. (Mis-) trust in governments and healthcare quality has a significant (negative) positive impact on vaccination.
5. Finally, the estimated magnitude of the effects found is in line with the literature; for instance, in the US, Kreps, et al. (2020) estimated that being exposed to a virus from known others, increases the will to get vaccinated by about 5%. In another study by Karlsson and colleagues (2021), it is estimated that increased perception of the general health severity of the covid-19 boosts vaccine intent between 12% to 25%.

Table 4. High level effects on covid-19 vaccine intent

Effects	Marginal effects on vaccine intent
---------	------------------------------------

Media perception (exaggerate the situation)	-13.10%
Infection risk perception and health worry	21.80%
Medical help appropriateness	10.10%
Government trust	12.30%
NPI compliance negative side effects	10.30%
Not exposed to covid	-13.20%

## Machine learning models

Even if already informative, traditional regression techniques may have important caveats. As an example, the regression model may have endogeneity issue; if media information can affect vaccine intent, vaccine intent can also affect how people absorb media information, as a reverse causality effect. Likewise, omitted variables might be a challenge as e.g., non-measured effects such as vaccine effectiveness, can both influence vaccine and NPI intent. Finally, a lot of the effects can be nonlinear, e.g. trust in authorities may influence risk perceptions, but more or less depending on gender, age or income (see Wachinger et al., 2013).

We thus resort to deep learning approaches, that allow for providing more robust predictions given the restrictions of the above, but also because we want to analyze non-linear effects, that could be of high importance to reveal nudges to reduce the portion of refusers and hesitant. In particular, we have tested multiple learning techniques, from GVM to neural networks, etc. As we have found that the techniques provide same qualitative results, we herewith report on the Random Forest (RF) technique, as the technique is the most often used as big data technique related to pandemics; for example, Random Forest techniques have exhibited superior predictive power for H5N1 influenza outbreaks see Kane et al. (2014), and recently for covid-19 infections (Yadav et al., 2020; Yeşilkanat, 2020). Here, we have configured the Random Forest algorithm with 5,000 trees in the forest, including statements and demographics, with all variables being tried at each split, in order to understand hidden complex tree structure to predict vaccination intent. The model was run for each country specifically and we present the results by country in order to gauge country commonalities and specificities. The analysis also looks at the three key segments of vaccine preferences (acceptors (YES), refusers (NO) and hesitant (HST)). Table 5 lays out the predictive power of the RF algorithm per country based on root mean square error (RMSE). Tables 6 to 10 provide results for each of the five countries in terms of contribution to accuracy (in points of percentage), and by segments of vaccine preferences.

If the RMSE remains large, it is nevertheless in the range of 0.25-0.5 for RF models. This level of accuracy supports the fact that the model can relatively predict the data accurately. Further, we have observed that the RMSE of RF techniques is lower than with the logistic regression. This is especially the case for hesitant and refusers, --two segments that are problematic, and must be nudged to change their mind if one wants to ensure herd immunity.

*Table 5. Predictive performance comparison (10-fold cross-validation) of Random Forest models*

Model	YES	HST	NO
	<b>RMSE</b>		
Germany	0.4205	0.3954	0.3299
France	0.4581	0.4374	0.3492
Italy	0.4223	0.3804	0.2886
Spain	0.4158	0.3720	0.2692
Sweden	0.4636	0.4380	0.3511
Model	YES	HST	NO
	<b>RMSE RF &gt; RMSE logit</b>		
Germany	YES		YES
France		YES	YES
Italy		YES	YES
Spain		YES	
Sweden	YES	YES	

One confirms, from those tables, that RF generates a broader list of variables (from 31 for Spain to up to 41 in France) that affect vaccine preferences, than the list of 17 variables that has emerged out of traditional techniques. We also note that the difference seems to be that RF provides a longer-tail of variables that are often either country, or segment specific.

### Common factors

Regarding common factors, two elements converge with the logistic model predictions. Health risk perceptions are largest contributors to vaccine preferences prediction accuracy. “Being worried about own health”, and “coronavirus is dangerous for my health”, clearly discriminate between acceptor of vaccine or not, in all countries. Indirect perception linked to media further shapes the type of non-acceptance, as media exaggeration is a consistent tag line for refusers (versus hesitant), while media reliability of information emerges consistently across countries as a universal driver of acceptance (versus hesitation).

We also witness the critical importance of institutional trust, a factor that has been noted as having one of the most substantial impacts on risk perception of natural hazards; see Yadav et al., 2020, for the covid-19 as well as Terpstra (2011) for flood impact. Trust (as well as satisfaction) in healthcare/government (actions) both play a role, with the trust in healthcare generally slightly more predictive for vaccination, than trust in government actions.

On top of the logistic regression, RF provides additional elements that were hidden by the technique of regression, both because they affect segment preferences asymmetrically, and/or because the effects of those factors were absorbed and tied to other variables:

1. The first is peer trust, acting on top of institutional trust. This *peer trust* is measured by two markers, « covid reveals the best/worst of people », and « The Covid crisis will bring countries closer ». The lack of peer trust makes vaccination less likely, as it was also shown to reduce NPI compliance during the covid-19, see Mehari (2020). As peer trust is correlated with government trust, the effect got blurred into government trust in the regression analysis;
2. The second element is the *balance between economics and health*, as preference for the later versus the former leads to a clear separation between refusers and acceptors of vaccine uptake. This effect has remained hidden in the regression technique, as overweight to health correlates strongly with health worries, while the effect on vaccination depends as well on other elements such as age.
3. The third is that, as per the conceptual model, *compliance with NPI measures predicts vaccine intent*, while indirect side effects of those measures, especially bad-being linked to isolation, also leads to more vaccine intent. Note that the isolation effect is large enough to have refusers switch to become more hesitant, but still not yet fully convinced to be vaccinated.
4. The fourth element is *the expectation of the duration of the crisis*. An expectation that the crisis will not last long reduces acceptors in favor of hesitant for all countries under analysis.

Table 6. “Acceptors”, “Hesitant” and “Refusers”: Factors in common or country specific - Germany

Factors	#	YES	HST	NO
I actively encourage others to follow the restrictions and guidelines	3	17.131	1.344	6.477
I am worried about my own health	3	4.597	1.185	2.484
Coronavirus is dangerous for my health	3	20.196	1.376	8.585
Media exaggerate the situation with COVID-19	3	5.507	1.144	14.41
Media provide reliable information about the pandemic	3	2.976	1.183	1.578
The government is doing a good job dealing with COVID-19	3	4.752	1.319	1.799
# of children 0	3	2.747	1.254	1.571
I comply with the restrictions to stay home	2	3.107	1.92	
I wash hands for 20 seconds when necessary	2	2.586	2.495	
COVID-19 increases domestic violence	2	2.604		1.212
I am anxious about not being able to meet with friends	2	3.713		1.543

I am grateful to our healthcare professionals	2	4.457	3.449	
I am satisfied with how my government is handling this crisis	2	5.599	4.834	
I am satisfied with how our healthcare system is handling this crisis	2	2.728	2.157	
In case of a coronavirus infection I will get appropriate medical help	2	3.858		1.24
The government discloses real numbers of coronavirus infections and deaths	2	4.062		1.171
The restrictions caused by COVID-19 will continue for about a month	2	4.085		1.691
politics Other	2	3.148	1.507	
age 26-35	1		1.5	
I would like to help people who are more vulnerable to COVID-19	1			1.226
I am worried about my financial situation	1			1.147
I am worried that our country will run out of money	1		1.552	
I am worried that there will not be enough basic necessities in the stores	1		1.811	
I am worried about the health of my children	1		1.586	
I am worried about the health of my older family members	1		1.565	
I am worried about my children s education	1		1.75	
COVID-19 will increase divorce rates	1			1.137
I am worried about not being able to meet with my family	1			1.486
Living in isolation negatively impacts my wellbeing	1			1.322
COVID-19 will bring countries closer	1	2.683		
Slowing the spread of COVID-19 is more important than the economy	1			1.218
[PRESIDENT] is doing a good job dealing with COVID-19	1	2.63		
COVID-19 reveals the worse in people	1		2.259	
People will stop following the restrictions soon	1			1.174
The restrictions caused by COVID-19 will continue at least until the fall	1			1.131

Table 7. “Acceptors”, “Hesitant” and “Refusers”: Factors in common or country specific – France

Factors	#	H2T	NO	YES
I disinfect groceries before putting them away	3	1.51	3.015	6.175
In case of a coronavirus infection I will get appropriate medical help	3	1.426	2.601	4.301
Don't associate with politics	3	1.881	1.866	5.621
I wash hands for 20 seconds when necessary	2		2.558	3.296
I would like to help people who are more vulnerable to COVID-19	2	1.517	1.89	
Since COVID-19 I exercise less	2	1.532		3.313
I am worried about my own health	2		3.089	6.441
Being together all the time increases family tensions	2	1.535	1.865	
I am anxious about not being able to meet with friends	2	1.429		3.486
Coronavirus is dangerous for my health	2		3.328	5.738
Media exaggerate the situation with COVID-19	2	1.429	3.785	
I am satisfied with how my government is handling this crisis	2	1.527		4.175
I am satisfied with how our healthcare system is handling this crisis	2	1.581		3.499
The government discloses real numbers of coronavirus infections and deaths	2	1.444		3.574
People will stop following the restrictions soon	2	1.478	1.89	
The restrictions caused by COVID-19 will continue at least until the fall	2	1.711		4.218
age 36-49	1		2.354	
age 50-64	1		1.718	
age >64	1			3.977
I disinfect mail and deliveries before opening them	1			4.528
Since COVID-19 I eat more healthy	1	1.574		
Since COVID-19 I exercise at home more	1			3.071
I am worried about my financial situation	1	1.487		
I am worried about the health of my older family members	1		1.909	
COVID-19 will increase divorce rates	1	1.597		
I worry how living in isolation will affect me	1	1.597		
Living in isolation negatively impacts my wellbeing	1	1.523		

COVID-19 will bring countries closer	1			6.177
I am grateful to our essential workers	1		1.848	
Media provide reliable information about the pandemic	1			4.336
COVID-19 reveals the best in people	1			3.578
COVID-19 reveals the worse in people	1		1.904	
I believe we will beat COVID-19 soon	1		2.028	
The restrictions caused by COVID-19 will continue for about a month	1	1.562		
Middle school	1		1.673	
Vocational	1		1.924	
income <20000€	1		1.961	
Retired	1			5.051
politics Right	1			3.577
politics Other	1		1.81	
Male	1	1.456		

Table 8. “Acceptors”, “Hesitant” and “Refusers”: Factors in common or country specific – Italy

Factors	#	H2T	NO	YES
Since COVID-19 I exercise at home more	3	1.254	1.023	2.883
I am worried about my own health	3	1.051	1.168	3.03
Coronavirus is dangerous for my health	3	1.694	3.883	15.414
Media provide reliable information about the pandemic	3	1.317	1.433	5.733
In case of a coronavirus infection I will get appropriate medical help	3	1.553	1.369	5.343
income <20000€	3	1.439	1.493	2.67
age 36-49	2		1.147	3.072
I actively encourage others to follow the restrictions and guidelines	2		4.869	4.075
Since COVID-19 I exercise less	2	1.056	1.026	
I am worried about the health of my older family members	2	1.063		2.595
Slowing the spread of COVID-19 is more important than the economy	2	1.218		2.816
[PRESIDENT] is doing a good job dealing with COVID-19	2	1.37		3.08
The government is doing a good job dealing with COVID-19	2	1.246		3.339
The government discloses real numbers of coronavirus infections and deaths	2	1.109		4.001
I believe we will beat COVID-19 soon	2		1.864	2.422
income >20000€	2	1.544		2.451
income don't want to answer	2		1.348	2.566
I comply with the restrictions to stay home	1		1.039	
I disinfect groceries before putting them away	1		1.02	
I wash hands for 20 seconds when necessary	1		1.455	
I would like to help people who are more vulnerable to COVID-19	1	1.131		
Since COVID-19 I eat more healthy	1	1.106		
Since COVID-19 I eat more unhealthy	1		1.073	
I am worried that our country will run out of money	1		1.548	
Being together all the time increases family tensions	1			2.679
COVID-19 will increase divorce rates	1		1.146	
I am worried about not being able to meet with my family	1	1.049		
I worry how living in isolation will affect me	1	1.191		
Living in isolation negatively impacts my wellbeing	1	1.241		
Media exaggerate the situation with COVID-19	1		1.259	
I am satisfied with how our healthcare system is handling this crisis	1	1.125		
Vocational	1		1.534	
High school	1			2.424
# of children 2	1			2.601
politics Left	1			3.113
politics Other	1	1.094		
Don't associate with politics	1		1.029	

Table 9. “Acceptors”, “Hesitant” and “Refusers”: Factors in common or country specific – Spain

Factors	#	H2T	NO	YES
---------	---	-----	----	-----

I would like to help people who are more vulnerable to COVID-19	3	1.067	0.5574	2.967
I am worried that there will not be enough basic necessities in the stores	3	1.172	0.6056	3.035
I am worried about my own health	3	1.245	0.5962	3.148
Being together all the time increases family tensions	3	1.398	0.6332	4.277
I am worried about not being able to meet with my family	3	1.061	0.679	3.564
I worry how living in isolation will affect me	3	1.071	0.6362	2.667
Slowing the spread of COVID-19 is more important than the economy	3	1.01	0.6841	3.138
Coronavirus is dangerous for my health	3	1.236	0.6955	4.37
I am satisfied with how our healthcare system is handling this crisis	3	1.19	0.5885	3.314
Don't associate with politics	3	1.316	0.6062	5.168
Since COVID-19 I eat more healthy	2	1.079		2.739
I worry that there will be an increase in break-ins and thefts	2	1.067		2.759
I am grateful to our essential workers	2		0.9236	5.093
I am grateful to our healthcare professionals	2		0.8812	3.935
Media exaggerate the situation with COVID-19	2		1.0383	3.069
COVID-19 reveals the best in people	2	1.179		2.525
COVID-19 reveals the worse in people	2	1.032	0.581	
The restrictions caused by COVID-19 will continue at least until the fall	2		0.6249	2.762
The restrictions caused by COVID-19 will continue for about a month	2	1.21		2.515
I disinfect groceries before putting them away	1		0.6347	
I disinfect mail and deliveries before opening them	1		0.7	
Since COVID-19 I exercise less	1	1.015		
The COVID-19 outbreak will make society more unequal	1	1.048		
I am worried about the health of my older family members	1		0.613	
I am anxious about not being able to meet with friends	1		0.6025	
Living in isolation negatively impacts my wellbeing	1	1.066		
COVID-19 will bring countries closer	1	1.02		
My chance of getting COVID-19 is high	1			2.676
In case of a coronavirus infection I will get appropriate medical help	1		0.5866	
People will stop following the restrictions soon	1			2.813
income >20000€	1	1.004		

Table 10. "Acceptors", "Hesitant" and "Refusers": Factors in common or country specific - Sweden

Factors	#	H2T	NO	YES
I would like to help people who are more vulnerable to COVID-19	3	1.628	0.9341	3.305
I am worried about my own health	3	1.521	1.1974	7.91
In case of a coronavirus infection I will get appropriate medical help	3	1.676	1.0214	6.545
The government discloses real numbers of coronavirus infections and deaths	3	1.498	1.0189	3.421
People will stop following the restrictions soon	3	1.595	1.0771	3.231
Since COVID-19 I exercise at home more	2		0.9342	2.947
I am worried about my financial situation	2	1.474	0.9647	
I am worried that our country will run out of money	2	1.534	1.0807	
Living in isolation negatively impacts my wellbeing	2	1.488		3.331
Slowing the spread of COVID-19 is more important than the economy	2		1.0827	3.512
Coronavirus is dangerous for my health	2		1.289	5.605
Media provide reliable information about the pandemic	2	1.506		3.954
[PRESIDENT] is doing a good job dealing with COVID-19	2	1.575		4.894
I am satisfied with how my government is handling this crisis	2		0.909	3.079
The government is doing a good job dealing with COVID-19	2	1.581		3.712
I am satisfied with how our healthcare system is handling this crisis	2	1.7		3.264
COVID-19 reveals the best in people	2	1.491		3.67
COVID-19 reveals the worse in people	2	1.547	0.909	
Quarantine no	2	1.52		3.218
age 26-35	1		1.0788	
age 36-49	1		1.1173	
age >64	1			6.116

I comply with the recommendations for physical distancing	1		0.9211	
I comply with the restrictions to stay home	1		1.0453	
The COVID-19 outbreak will make society more unequal	1		0.9766	
I am worried about the health of my older family members	1		1.0895	
COVID-19 increases domestic violence	1			3.007
COVID-19 will increase divorce rates	1	1.55		
I am worried about not being able to meet with my family	1		1.117	
COVID-19 will bring countries closer	1			4.05
My chance of getting COVID-19 is high	1			3.176
Media exaggerate the situation with COVID-19	1		1.5861	
I believe we will beat COVID-19 soon	1	1.477		
The restrictions caused by COVID-19 will continue at least until the fall	1	1.483		
The restrictions caused by COVID-19 will continue for about a month	1	1.553		
Don't associate with politics	1	1.469		

## Specific factors

From the list, a few specific factors finally emerge that are either country or segment dependent.

In general country differences can be expected, and in particular because of the different measures undertaken by the authorities. (Full country comparison can be found in Tables 11, 12 and 13 in appendix) For example, while third parties' health worries are positively associated with higher propensity to get the vaccine, third party's most concern is about children in Germany, while it is more about older family members in Italy. Italy in fact, was put in a stricter lockdown than Germany, due to pandemic not being in control, and large mortality happening among older population. This lockdown made the Italians more at risk to eat unhealthy and not exercise; those able to stick to healthy practices were thus less inclined to feel bad about lockdown, with higher reluctance to vaccinate. Isolation and inability to meet friends were not seen as tangible risks in Sweden given that no strict lockdown was prevailing in the first wave of the pandemic. Lockdown in form of quarantine when sick was the only hazard Swedish citizens had to face, as social distancing was not mandatory for citizens; thus being confronted with being contaminated was a way to reduce anti-vaccination in the country. Segment differences are more visible with RF than with traditional regressions, in general, because the effects play non-linearly. Age does not play a role for hesitant, but more between acceptors and those reluctant to get vaccinated, and especially for all age brackets in France. Low income plays a negative role on intent, especially for Italy, as an example.

In general and as the various Tables 6-10 show, the extra factors emerging from the RF exercise (about 20 more factors) are essentially effects that are clustering by segment preferences and/or countries, and are typically known correlates of other main factors (e.g. low income, low education correlate with lower citizen trust).

We conclude by looking at the hesitant, as this segment is twice as large as the refusers, and is the one that one would hope to influence to convert to vaccination as a way to reach herd immunity. Logit regression results shown in Table 3 suggest that hesitant are a category of less assertive refusers, except that family tension (but not violence) leads them to convert to accept vaccine. What RF does show is that you can find better nudges to make them convert. First, this is linked to the reality of the pandemic, that is pandemic will not stop easily and will last long without enough vaccinated to reach immunity. Second, it is important to demonstrate better institutional trust, with much more appropriate and educated media information. This effect of trust has been already shown to have a substantial effect on risk perception in many natural hazards. We add covid-19 pandemic to this list.

## CONCLUSIONS

This article has used big data techniques, in particular Random Forest technique, to provide a more robust baseline regarding the covid-19 vaccine preferences. As we hypothesized that when it comes to vaccine decisions, a large set of factors interact and possibly with asymmetry, we demonstrate that RF provides a much more granular view as to the factors affecting vaccine choices. Good news is that the

techniques do not necessarily conflict. The same common, first-order factors emerge as predictors in all techniques, but RF clearly provides a more complete view.

In general, the research is a first to show that compliance to non-pharma protective measures is already a good marker of the citizen mindset for vaccine acceptance, as is the perception of how long will the crises last. As per other literature, the key nudge remains to objective the real risk of the pandemic, and make people also aware of the fact that a pandemic is only to be stopped by limiting self-defeating behavior, like reluctance to, or free-riding on protective measures.

## ACKNOWLEDGMENT

Special acknowledgement should be given to NEUROHM for initiating and supervising the projects and SYNO for collecting the data.

## REFERENCES

- Amatriain, X. (2013) *Beyond Data: from user information to business value through personalized recommendations and consumer science*. Paper presented at the CIKM'13. San Francisco, CA.
- Atwell, K., Lake, J., Sneddon, J., Gerrans, P., Blyth, C., & Lee, J. (2021). Converting the maybes: Crucial for a successful COVID-19 vaccination strategy. *PLoS ONE*, 16(1).
- Bargh J. A. (1997). Automaticity of Everyday Life. *Advances in Social Cognition*, 10, 1-61.
- Bauer, P., Stevens, B., & Hazeleger, W. (2021). A digital twin of Earth for the green transition. *Nature Climate Change*, 11(2), 80-83.
- Bornstein R. F., & Pittman T. S. (1992). *Perception Without Awareness*. New York: The Guilford Press.
- Becker, M. H. (1974). The health belief model and sick role behavior. *Health education monographs*, 2, 409-419.
- Brewer, N. T., Chapman, G. B., Gibbons, F. X., Gerrard, M., McCaul, K. D., & Weinstein, N. D. (2007). Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. *Health Psychology*, 26, 136.
- Brynjolfsson, E., Hitt, L., & Kim, H. (2011) Strength in numbers: how does data-driven decision making affect firm performance?. Available at: SSRN - <https://ssrn.com/abstract=1819486>
- Bughin, J. (2016). Big data, Big bang? *Journal of Big Data*, vol. 3, no 1, p. 1-14.
- Bughin, J., & Cincera, M. (2021). How Institutional Actions before Vaccine Affects Vaccination Intention Later: Prediction Via Machine Learning. Available at: SSRN - <https://ssrn.com/abstract=3824550>
- Bughin, J., Cincera, M., Reykowska, D., Żyszkiewicz, M., & Ohme, R. (2021). COVID-19 Endemism and the Control Skeptics. Available at: SSRN - <https://ssrn.com/abstract=3785230>
- Bughin, J., Cincera, M., Reykowska, D., Żyszkiewicz, M., & Ohme, R. (2020). Perceptive risk clusters of European citizens and NPI compliance in face of the covid-19 pandemics. *Working Papers ECARES 2020-51*, ULB Université libre de Bruxelles.
- Cho, S. W. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. *The Econometrics Journal*, 23(3), 323-34.
- Eissler, M. (2021). Probability that an infection like Covid-19 stops without reaching herd immunity, calculated with a stochastic agent-based model. *medRxiv*, 2020-12.
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, 15, 115-141.
- Fazio, R. H., Powell, M. C., & Williams, C. J. (1989). The role of attitude accessibility in the attitude-to-behavior process. *Journal of Consumer Research*, 16(3), 280-288.



- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of personality and social psychology*, 50(2), 229.
- Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude–perception and attitude–behavior relations: An investigation of the 1984 presidential election. *Journal of personality and social psychology*, 51(3), 505.
- Greenwald A. G. (1992). New Look 3: Unconscious cognition reclaimed, *American Psychologist*, 47, 766-779.
- Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC bioinformatics*, 15(1), 276.
- Karlsson, L. C., Soveri, A., Lewandowsky, S., Karlsson, L., Karlsson, H., Nolvi, S., Karuki, M., Lindfelt, M., & Antfolk, J. (2021). Fearing the disease or the vaccine: The case of COVID-19. *Personality and individual differences*, 172, 110590
- Kraus, S. J. (1995). Attitudes and the Prediction of Behavior: A Meta-Analysis of the Empirical Literature. *Personality and Social Psychology Bulletin*, 21(1), 58–75.
- Kreps, S., Prasad, S., Brownstein, J. S., Hswen, Y., Garibaldi, B. T., Zhang, B., & Kriner, D. L. (2020). Factors associated with US adults' likelihood of accepting COVID-19 vaccination. *JAMA network open*, 3(10), e2025594-e2025594.
- Larson, H., Jarrett, C., Eckersberger, E., Smith, D., & Paterson, P. (2014). Understanding vaccine hesitancy around vaccines from a global perspective : a systematic review of published literature, 2007-2012, *Vaccine* (32): 2150-59.
- Mavragani, A., & Tsagarakis, K. P. (2019). Predicting referendum results in the Big Data Era. *Journal of Big Data*, 6(1), 1-20
- Mehari, Y. (2020). The Role of Social Trust in Citizen Mobility During COVID-19. Available at: SSRN - <https://ssrn.com/abstract=3607668>.
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *Journal of personality and social psychology*, 64(5), 723.
- Nguyen, A., & Catalan, D. (2020). Digital mis/disinformation and public engagement with health and science controversies: fresh perspectives from Covid-19. *Media and Communication*, 8(2), 323-328.
- Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in cognitive sciences*, 15(4), 152-159.
- Ohme, R. (2001). The implicit conditioning of consumer attitudes: Logo substitution effect. In R. Ohme (Ed.) *Implicit Processes in Cognition*. Special Edition of Polish Psychological Bulletin, 32(1).
- Ohme, R., Matukin, M., & Wicher, P. (2020). Merging Explicit Declarations With Implicit Response Time to Better Predict Behavior. In V. Chkoniya, A. O. Madsen & P. Bukhrashvili (Ed.), *Anthropological Approaches to Understanding Consumption Patterns and Consumer Behavior* (pp. 427-448). IGI Global.
- Ruiz, J. B., & Bell, R. A. (2021). Predictors of intention to vaccinate against COVID-19: Results of a nationwide survey. *Vaccine*, 39(7), 1080-1086.
- Sallam, M. (2021). COVID-19 vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates. *Vaccines*, 9(2), 160.
- Terpstra, T. (2011). Emotions, trust, and perceived risk: affective and cognitive routes to flood preparedness. *Risk Analysis*, 31(10), 1658-1675.

- Uleman J. S. & Bargh J. A. (1989). *Unintended thought*. New York: Guilford Press.
- Wachinger, G., Renn, O., Begg, C., & Kuhlicke, C. (2013). The risk perception paradox: Implications for governance and communication of natural hazards. *Risk Analysis*, 33(6), 1049–1065.
- Yadav, M., Perumal, M., & Srinivas, M. (2020). Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos, Solitons & Fractals*, 139, 110050.
- Yang, C. , Dillard, J. , & Li, R. (2018). Understanding fear of Zika: Personal, interpersonal, and media Influences. *Risk Analysis*, 38(12), 2535–2545
- Yeşilkanat, C. M. (2020). Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine-learning algorithm. *Chaos, Solitons & Fractals* (140): 110210.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9, 1-27.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American psychologist*, 35(2), 151.
- Zickfeld, J. H., Schubert, T. W., Herting, A. K., Grahe, J., & Faasse, K. (2020). Correlates of health-protective behavior during the initial days of the COVID-19 outbreak in Norway. *Frontiers in psychology*, 11.

## ADDITIONAL READING

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Jovančević, A., & Milićević, N. (2020). Optimism-pessimism, conspiracy theories and general trust as factors contributing to COVID-19 related behavior—A cross-cultural study. *Personality and individual differences*, 167, 110216.
- Lazarus, J. V., Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., & El-Mohandes, A. (2020). A global survey of potential acceptance of a COVID-19 vaccine. *Nature medicine*, 1-4.
- Neumann-Bohme, S., Varghese, N. E., Sabat, I., Barros, P. P., Brouwer, W., Van Exel, J., & Stargardt T. (2020). Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. *The European Journal of Health Economics* (21): 977–982.

## KEY TERMS AND DEFINITION

**Attitudes:** Can be defined as an organized predisposition to respond in a favorable or unfavorable manner toward a specified class of objects.

**Big data:** Big and versatile data sets, that are hard to analyze using standard analytical methods. Analysis of big data enables to identify new information and gain knowledge by finding relationships between variables not visible before.

**iCode™:** A unique tool created by NEUROHM to test response time (RT) in the field of brands, products, commercials, motives, emotional experience, sensory testing, services, satisfaction or personal assessment.

**Machine learning:** An area of artificial intelligence dedicated to algorithms that improve automatically through experience obtained via data exposure. Machine learning algorithms

build a mathematical model from sample data, called a training set, to forecast or make decisions without being explicitly programmed by humans for that purpose. Machine learning algorithms are used in many different applications, such as image recognition, speech recognition, data analysis and classification, anti-spam etc.

**Random Forest:** Random forest is a supervised learning algorithm that works as an ensemble learning method based on building a range of regression trees that are then averaged out to compose the final forest. From each tree, the smallest root mean square error determines the top of the tree and recursively creates a full tree. Prediction uses the average of the response variable in each leaf of the tree. The advantages of Random Forests are: they are robust to outliers, effective with nonlinear data, and have lower risk of overfitting compared to decision trees.

**Response time (RT):** Time necessary to produce an answer. Method frequently used by psychologists to measure e.g. attitude strength.

## **APPENDIX**

Table 11. Factors in common and specific to countries: “Acceptors”

Factors	#	DE	FR	IT	SE	SP
I am worried about my own health	5	1	1	1	1	1
Coronavirus is dangerous for my health	5	1	1	1	1	1
Media provide reliable information about the pandemic	4	1	1	1	1	
I am satisfied with how our healthcare system is handling this crisis	4	1	1		1	1
In case of a coronavirus infection I will get appropriate medical help	4	1	1	1	1	
The government discloses real numbers of coronavirus infections and deaths	4	1	1	1	1	
Since COVID-19 I exercise at home more	3		1	1	1	
COVID-19 will bring countries closer	3	1	1		1	
Slowing the spread of COVID-19 is more important than the economy	3			1	1	1
[PRESIDENT] is doing a good job dealing with COVID-19	3	1		1	1	
I am satisfied with how my government is handling this crisis	3	1	1		1	
The government is doing a good job dealing with COVID-19	3	1		1	1	
COVID-19 reveals the best in people	3		1		1	1
age >64	2		1		1	
I actively encourage others to follow the restrictions and guidelines	2	1		1		
I wash hands for 20 seconds when necessary	2	1	1			
I would like to help people who are more vulnerable to COVID-19	2				1	1
Being together all the time increases family tensions	2			1		1
COVID-19 increases domestic violence	2	1			1	
I am anxious about not being able to meet with friends	2	1	1			
I am grateful to our healthcare professionals	2	1				1
My chance of getting COVID-19 is high	2				1	1
Media exaggerate the situation with COVID-19	2	1				1
People will stop following the restrictions soon	2				1	1
The restrictions caused by COVID-19 will continue at least until the fall	2		1			1
The restrictions caused by COVID-19 will continue for about a month	2	1				1
Don't associate with politics	2		1			1
I comply with the restrictions to stay home	1	1				
I disinfect groceries before putting them away	1		1			
I disinfect mail and deliveries before opening them	1		1			
Since COVID-19 I eat more healthy	1					1
Since COVID-19 I exercise less	1		1			
I am worried that there will not be enough basic necessities in the stores	1					1
I am worried about the health of my older family members	1			1		
I worry that there will be an increase in break-ins and thefts	1					1
I am worried about not being able to meet with my family	1					1
I worry how living in isolation will affect me	1					1
Living in isolation negatively impacts my wellbeing	1				1	
I am grateful to our essential workers	1					1
I believe we will beat COVID-19 soon	1			1		
High school	1			1		
income <20000€	1			1		
income >20000€	1			1		
income don't want to answer	1			1		
# of children 0	1	1				
# of children 2	1			1		
Retired	1		1			
politics Left	1			1		
politics Right	1		1			
politics Other	1	1				
Quarantine no	1				1	

Table 12. Factors in common and specific to countries: “Hesitant”

Factors	#	DE	FR	IT	SE	SP
I would like to help people who are more vulnerable to COVID-19	5	1	1	1	1	1
Living in isolation negatively impacts my wellbeing	5	1	1	1	1	1
I am worried about my own health	4	1		1	1	1
I am satisfied with how our healthcare system is handling this crisis	4		1	1	1	1
In case of a coronavirus infection I will get appropriate medical help	4	1	1	1	1	
The government discloses real numbers of coronavirus infections and deaths	4	1	1	1	1	
The restrictions caused by COVID-19 will continue for about a month	4	1	1		1	1
Since COVID-19 I eat more healthy	3		1	1		1
Since COVID-19 I exercise less	3		1	1		1
I am worried about my financial situation	3	1	1		1	
COVID-19 will increase divorce rates	3	1	1		1	
I am worried about not being able to meet with my family	3	1		1		1
I worry how living in isolation will affect me	3		1	1		1
Slowing the spread of COVID-19 is more important than the economy	3	1		1		1
Coronavirus is dangerous for my health	3	1		1		1
Media provide reliable information about the pandemic	3	1		1	1	
The government is doing a good job dealing with COVID-19	3	1		1	1	
People will stop following the restrictions soon	3	1	1		1	
The restrictions caused by COVID-19 will continue at least until the fall	3	1	1		1	
Don't associate with politics	3		1		1	1
Being together all the time increases family tensions	2		1			1
I am anxious about not being able to meet with friends	2	1	1			
Media exaggerate the situation with COVID-19	2	1	1			
COVID-19 reveals the best in people	2				1	1
COVID-19 reveals the worse in people	2				1	1
income >20000€	2			1		1
I actively encourage others to follow the restrictions and guidelines	1	1				
I disinfect groceries before putting them away	1		1			
Since COVID-19 I exercise at home more	1			1		
I am worried that our country will run out of money	1				1	
I am worried that there will not be enough basic necessities in the stores	1					1
The COVID-19 outbreak will make society more unequal	1					1
I am worried about the health of my older family members	1			1		
I worry that there will be an increase in break-ins and thefts	1					1
COVID-19 increases domestic violence	1	1				
COVID-19 will bring countries closer	1					1
[PRESIDENT] is doing a good job dealing with COVID-19	1			1		
[PRESIDENT] is doing a good job dealing with COVID-19	1				1	
I am satisfied with how my government is handling this crisis	1		1			
I believe we will beat COVID-19 soon	1				1	
income <20000€	1			1		
# of children 0	1	1				
politics Other	1			1		
Quarantine no	1				1	
Male	1		1			

Table 13. Factors in common and specific to country: “Refusers”

Factors	#	DE	FR	IT	SE	SP
I am worried about my own health	5	1	1	1	1	1
Coronavirus is dangerous for my health	5	1	1	1	1	1
Media exaggerate the situation with COVID-19	5	1	1	1	1	1
I am worried about the health of my older family members	4	1	1		1	1
In case of a coronavirus infection I will get appropriate medical help	4		1	1	1	1
COVID-19 reveals the worse in people	4	1	1		1	1
age 36-49	3		1	1	1	
I comply with the restrictions to stay home	3	1		1	1	
I disinfect groceries before putting them away	3		1	1		1
I wash hands for 20 seconds when necessary	3	1	1	1		
I would like to help people who are more vulnerable to COVID-19	3		1		1	1
I am worried that our country will run out of money	3	1		1	1	
Don't associate with politics	3		1	1		1
age 26-35	2	1			1	
I actively encourage others to follow the restrictions and guidelines	2	1		1		
Since COVID-19 I exercise at home more	2			1	1	
I am worried that there will not be enough basic necessities in the stores	2	1				1
Being together all the time increases family tensions	2		1			1
I am worried about not being able to meet with my family	2				1	1
I am grateful to our essential workers	2		1			1
I am grateful to our healthcare professionals	2	1				1
Slowing the spread of COVID-19 is more important than the economy	2				1	1
Media provide reliable information about the pandemic	2	1		1		
I am satisfied with how my government is handling this crisis	2	1			1	
I am satisfied with how our healthcare system is handling this crisis	2	1				1
I believe we will beat COVID-19 soon	2		1	1		
People will stop following the restrictions soon	2		1		1	
Vocational	2		1	1		
income <20000€	2		1	1		
politics Other	2	1	1			
age 50-64	1		1			
I comply with the recommendations for physical distancing	1				1	
I disinfect mail and deliveries before opening them	1					1
Since COVID-19 I eat more unhealthy	1			1		
Since COVID-19 I exercise less	1			1		
I am worried about my financial situation	1				1	
The COVID-19 outbreak will make society more unequal	1				1	
I am worried about the health of my children	1	1				
I am worried about my children s education	1	1				
COVID-19 will increase divorce rates	1			1		
I am anxious about not being able to meet with friends	1					1
I worry how living in isolation will affect me	1					1
The government is doing a good job dealing with COVID-19	1	1				
The government discloses real numbers of coronavirus infections and deaths	1				1	
The restrictions caused by COVID-19 will continue at least until the fall	1					1
Middle school	1		1			
income don't want to answer	1			1		
# of children 0	1	1				



## WORKING PAPERS 2013

- 001** - Exploring europe's r&d deficit relative to the us: differences in the rates of return to r&d of young leading r&d firms - Michele Cincera and Reinhilde Veugelers
- 002** - Governance typology of universities' technology transfer processes - A. Schoen, B. van Pottelsberghe de la Potterie, J. Henkel.
- 003** - Academic Patenting in Belgium: Methodology and Evidence – M. Mejer.
- 004** - The impact of knowledge diversity on inventive performance at European universities – M. Mejer
- 005** - Cross-Functional Knowledge Integration, Patenting and Firm's Performance – M. Ceccagnoli, N. van Zeebroeck and R. Venturini.
- 006** - Corporate Science, Innovation and Firm Value, M. Simeth and M. Cincera



## WORKING PAPERS 2014

- 007** - Determinants of Research Production at top US Universities – Q. David.
- 008** - R&D financing constraints of young and old innovation leaders in the EU and the US – M. Cincera, J. Ravet and R. Veugelers
- 009** - Globalization of Innovation Production; A Patent-Based Industry Analysis – J. Danguy
- 010** - Who collaborates with whom: the role of technological distance in international innovation – J. Danguy



## WORKING PAPERS 2015

- 011** - Languages, Fees and the International Scope of Patenting – D. Harhoff , K. Hoisl, B. van Pottelsberghe de la Potterie , C. Vandepuut
- 012** - How much does speed matter in the fixed to mobile broadband substitution in Europe? – M. Cincera, L. Dewulf, A. Estache
- 013** - VC financing and market growth – Interdependencies between technology-push and market-pull investments in the US solar industry – F. Schock, J. Mutl, F. Täube, P. von Flotow
- 014** - Optimal Openness Level and Economic Performance of Firms: Evidence from Belgian CIS Data – M. Cincera, P. De Clercq, T. Gillet
- 015** - Circular Causality of R&D and Export in EU countries – D. Çetin, M. Cincera.
- 016** - Innovation and Access to Finance – A Review of the Literature – M. Cincera, A. Santos.



## WORKING PAPERS 2016

**017** - Effectiveness of Government intervention in the SME sector: Evidence from the Brussels-Capital Region – G. E. Fombasso, M. Cincera.

**018** - A review of corporate R&D intensity decomposition – P. Moncada-Pastemò-Castello.

**019** - The laws of action and reaction: on determinants of patent disputes in European chemical and drug industries – R. Kapoor, N. van Zeebroeck.

**020** - How do the normativity of headquarters and the knowledge autonomy of subsidiaries co-evolve? – M. Hansmans, G. Liu.



## WORKING PAPERS 2017

**021** - The case for offensive strategies in response to digital disruption – J. Bughin, N. van Zeebroeck.

**022** - Access to finance as a pressing problem: Evidence from innovative European firms – A. Santos, M. Cincera.

**023** - Platform play among incumbent firms: the wrong focus? – N. van Zeebroeck, J. Bughin.

**024** - Social Movements – M. Hansmans, K. van Bommel.

**025** - Decoding Patent Examination Services – L. Gimeno-Fabran, B. van Pottelsberghe de la Potterie.

**026** - Countries Attractiveness: An Analysis of EU Firms's Decisions to (De) Localize R&D Activities – M. Cincera, A. Santos.



## WORKING PAPERS 2018


**027** - The impact of EUREKA projects on the economic performance of R&D SMEs – M. Cincera, G. Fombasso.

**028** - Forecasting short-term transaction fees on a smart contracts platform – C. Hoffreumon, N. van Zeebroeck.

**029** - Sharing the risk of innovative investment: Assessing the effect of new European financing instrument - M. Cincera, A. Santos.

**030** - Academic scientists: The golden opportunity for high-tech companies – L. Dewulf, M. Cincera.



 **WORKING PAPERS 2019**

**031** - Experimentation, Learning and Stress. The Role of Digital Technologies in Strategy Change – J. Bughin, T. Kretschmer, N. van Zeebroeck.

**032** - Types of Innovation and Firm Performance – M. Cincera, E. Ince.

**033** - Competition and Innovation: Evidence from Worldwide Corporate R&D Spenders - M. Cincera, E. Ince, A. Santos.

 **WORKING PAPERS 2020**

**034** - Selectivity versus Reach: Flattening the Curve of Covid 19 for Joint Health and Economic Prosperity – J. Bughin.

**035** - How firms will affect the Future of Work – J. Bughin.

**036** - Making Sense of the Covid-19 (SARS-CoV-2) Pandemic – J. Bughin.

**037** - F.O.G. and Teleworking: Some Labor Economics of Covid-19 – J. Bughin , M. Cincera

**038** - Learning from 20 Years of Research on Innovation Economics – B. van Pottelsberghe de la Potterie.

**039** - The Value of Platform Strategy. It's the Ecosystem, Stupid! - J. Gossé, C. Hoffreumon, N. van Zeebroeck, J. Bughin.

**040** - Ten Moments of Truths for the Covid-19 Crisis – J. Bughin.

**041** – The Great Employee Divide: Clustering Employee « Well-being » Challenge during Covid-19 – J. Bughin, M. Cincera, D. Reykowska, M. Żyszkiewicz.

**042** - Perceptive Risk Clusters of European Citizens and NPI Compliance in face of the Covid-19 Pandemics – J. Bughin, M. Cincera, D. Reykowska, M. Żyszkiewicz, R. Ohme.

 **WORKING PAPERS 2021**

**043** - Make it or Break it: Vaccination Intention at the Time of Covid-19 – J. Bughin, M. Cincera, K. Peters, D. Reykowska, M. Żyszkiewicz, R. Ohme.

**044** - Covid-19 Endemism and the Control Skeptics - J. Bughin, M. Cincera, D. Reykowska, M. Żyszkiewicz, R. Ohme.

**045** - Vaccination or NPI? A Conjoint Analysis of German Citizens' Preferences – J. Bughin, M. Cincera, E. Kiepfer, D. Reykowska, F. Philippi, M. Żyszkiewicz, R. Ohme, D. Frank.

**046** - Is this Time Different? Corporate Resilience in the Age of Covid-19 – J. Bughin, S. Berjoan, F. Hinterman, Y. Xiong.

**047**- Big data is decision science: the case of covid-19 vaccination - J. Bughin, M. Cincera, D. Reykowska, R. Ohme.