Research Article

Do survey estimates of the public's compliance with COVID-19 regulations suffer from social desirability bias?

Martin Vinæs Larsen[*], Jacob Nyrup[†], Michael Bang Petersen[*]

Abstract: The COVID-19 pandemic has led governments to instate a large number of restrictions on and recommendations for citizens' behavior. One widely used tool for measuring compliance with these strictures are nationally representative surveys that ask citizens to self-report their behavior. But if respondents avoid disclosing socially undesirable behaviors, such as not complying with government strictures in a public health crisis, estimates of compliance will be biased upwards. To assess the magnitude of this problem, this study compares measures of compliance from direct questions to those estimated from list-experiments - a response technique that allows respondents to report illicit behaviors without individual-level detection. Implementing the list-experiment in two separate surveys of Danish citizens (n>5,000), we find no evidence that citizens under-report non-compliant behavior. We therefore conclude that survey estimates of compliance with COVID-19 regulations do not suffer from social desirability bias.

Supplements: Open data, Preregistered

A n epidemic is a particular form of crisis. In many societal crises, governments can ask citizens to "keep calm" and take the necessary steps to counter the crisis. This is not possible in an epidemic because when governments are dealing with an infectious disease, health outcomes are ultimately tied to citizens' behavior. It is the behavior of citizens--and, in particular, the physical distance they keep from each other--that ultimately drives the growth curve of an epidemic (Anderson, Heesterbeek, Klinkenberg, & Hollingsworth, 2020). Therefore, the COVID-19 epidemic has led governments to instate a large number of recommendations for citizens' behavior. Since these recommendations are hard to enforce, voluntary compliance by citizens is central (Johnson et al., 2020). In the parlance of public administration, citizens' co-production needs to be a key part of any government's strategy.

A large literature in Public Administration studies how governments can get citizens to co-produce public services (for recent reviews, see Voorberg, Bekkers, & Tummers, 2015; Pestoff, Brandsen, & Verschuere, 2013). It is imperative that governments take stock of this literature, as well as related literatures in science and political communication (e.g., Jamieson, Kahan, & Scheufele, 2017; Jørgensen, Bor, & Petersen, 2020; Pfattheicher, Nockur, Böhm, Sassenrath, & Petersen, 2020), when thinking about how to increase citizen compliance with government recommendations. However, no matter what tactics governments decide to use, they need to know whether citizens are diligent co-producers in the fight against COVID-19, because this will enable goverments to effectively evaluate these tactics and retain control of the epidemic.

* Aarhus University

† Oxford University

Address correspondence to Martin Vinæs Larsen at (mvl@ps.au.dk)

A way to do this, which has gotten a lot of attention in the current crisis, is the use of technological tracing of actual behaviors. For example, cell-phone companies are sharing data with governments in several countries, which can be used to monitor whether people are complying with curbs on movement (Pollina & Busvine, 2020), and private companies and governments are investing in the development of tracking apps. This type of data undoubtedly gives us some insight into whether citizens comply with government recommendations, but they also face key limitations. For one, there are both legal and ethical data protection concerns, which means that (democratic) governments will limit their use of them to avoid breaching privacy policies and/or to uphold citizens' institutional trust. There is also a class of behaviors that are harder to detect using these tools. While we can see how many people are leaving home each day and we can track congregation, we cannot observe how people act toward others if they meet. Do they hug and kiss friends they meet on the street? Do they shake hands? And when they come home, do they wash their hands? Such behaviors are key for countering an unfolding epidemic (Wimalawansa, 2020). From a democratic perspective, the best way to obtain information about such private behavior are with nationally representative surveys, where people self-report their behavior while keeping strict anonymity.

While normatively appealing, the reliance on survey self-reports will only provide authorities with unbiased estimates of compliance if respondents are honest. Unfortunately, respondents are not always honest. In particular, we know that respondents sometimes want to project an image of themselves as following prevailing norms and rules because they (potentially erroneously) believe that their survey answer could be revealed and that they might then face some social or even judicial sanction (see Blair et al., 2020). It is not clear whether survey questions regarding COVID-19 will create such a social desirability bias among those who do not comply. On the one hand, norms regarding social distancing are potentially quite strong, as they have implications for life and death. On the other hand, researchers can do much to signal to respondents that their responses will remain confidential, which should repress concerns that non-compliers will be sanctioned for telling the truth.

Social desirability bias can be uncovered using a list-experiment. Here respondents are presented with a list of behaviors, and are then asked to report how many of these behaviors they have engaged in. Respondents are randomly divided into two groups: the first receives a list of non-sensitive behaviors, and the second receives a list of the same behaviors as well as the sensitive behavior researchers are interested in. With a large enough sample, researches can estimate the proportion of people who have engaged in the sensitive behavior by looking at the average difference between the two groups. List-experiments help uncover social desirability bias by providing respondents with an additional level of anonymity, as the researcher can never infer whether a specific individual has engaged in the sensitive behavior (unless zero or all items are reported as true).

In this study, we compare estimates of compliance from such a list-experiment to estimates of compliance from direct questions, thereby identifying any potential social desirability bias. We do this by embedding a list-experiment into a rolling cross-sectional survey of Danish citizens' compliance with COVID-19 regulations and recommendations (n=3,515). Our design and analyses were pre-registered (see https://osf.io/ux6cs). As a follow-up we also did two non-preregistered list-experiments of sensitive prospective items in a separate survey (N=2,096). We find no evidence of underreporting in these list-experiments. Our results suggest that survey estimates of the public's compliance with COVID-19 regulations do not suffer from social desirability bias.

Study Design

We embedded a list-experiment in two separate surveys. The first is a rolling cross-sectional survey of Danish citizens that measures compliance with COVID-19 regulations and recommendations. The list-experiment was included in the survey from March 23 to March 30, 2020. 3,515 respondents participated in the survey in this period (i.e., n=3,515). The study was pre-registered on March 27. Data and results from the list-experiment were sent to the authors on March 31. The second survey included two list-experiments and was conducted from March 24 to April 1 (n=2,096). The survey company Epinion, who recruited respondents from their large, online panel, conducted both surveys.

In measuring compliance, we focus on the general recommendations from the Danish Health Authority regarding COVID-19.[1] In particular, the Danish Health Authorities instructed Danes to refrain from physical and in-person social contact. In our surveys, we operationalized this as whether respondents "hugged or kissed someone outside their immediate family", "attended a large social gathering", or "visited a friend or got a visit

from a friend". The Danish Health Authorities also recommended washing hands, being more diligent with cleaning, and coughing/sneezing into your sleeve, but we opted to focus on physical and in-person social contact, as these behaviors are easier to operationalize in a survey.

The first survey examined whether respondents reported "hugging or kissing someone outside their immediate family yesterday". Respondents were initially asked directly about whether they engaged in this behavior (they could answer Yes, No, or Don't know). The question came in a battery of other questions asking about compliance with other COVID-19 recommendations. The questions were prefaced by an appeal to answer honestly and highlighted that there were no right or wrong answers. Near the end of the survey, respondents were presented with the list-experiment. Respondents were randomly assigned to either a list of four non-sensitive behaviors or the same list plus the "hugging and kissing" item. By presenting respondents with both the direct questions and the list-experiment, rather than randomizing across which version was shown, we are able to compare respondents directly while maximizing statistical power. The list-experiment was placed at the end of the survey to avoid contaminating the rolling cross-sectional survey.

In the second survey, we make three changes to the basic design. First, we include the list-experiment before the direct question to assuage concerns related to question order effects. Second, instead of asking about past behavior, we ask about prospective behavior - what people plan to do. Third, we include the two items related to social contact: whether people were planning to "attend a large social gathering" or "visit a friend or get a visit from a friend". Appendix A presents an overview of the three list-experiments and Appendix B presents descriptive statistics for the two samples. Response distribution tables can be found in Appendix C.

Analysis: Comparing Estimates of Compliance

Figure 1 presents estimates of non-compliance with the government's advice from the list-experiment and the direct question. For the direct question we simply look at the proportion who said they would engage in the activity. For the list-experiment, compliance is estimated as the mean difference between those presented with the sensitive item and those who were not presented with this item. The standard errors are estimated using the bootstrapping method laid out in Blair et al. (2020). We use a similar method to estimate standard errors for the difference in compliance rates.[3] Overall, we see very low levels of non-compliance, and there are no systematic or significant differences between the direct questions and the list-experiments.

List-experiments are not very statistically efficient, so the confidence intervals are quite broad for each individual question, which means that our null finding could still be consistent with some under-reporting of non-compliance. To deal with this problem we also do a pooled analysis where we stack the three different questions and analyze them as one. When we do this, we get an almost identical estimate for non-compliance across direct question and list-experiment, and a 95 percent confidence interval allows us to rule out differences above 5.5 percentage points.

Figure 1
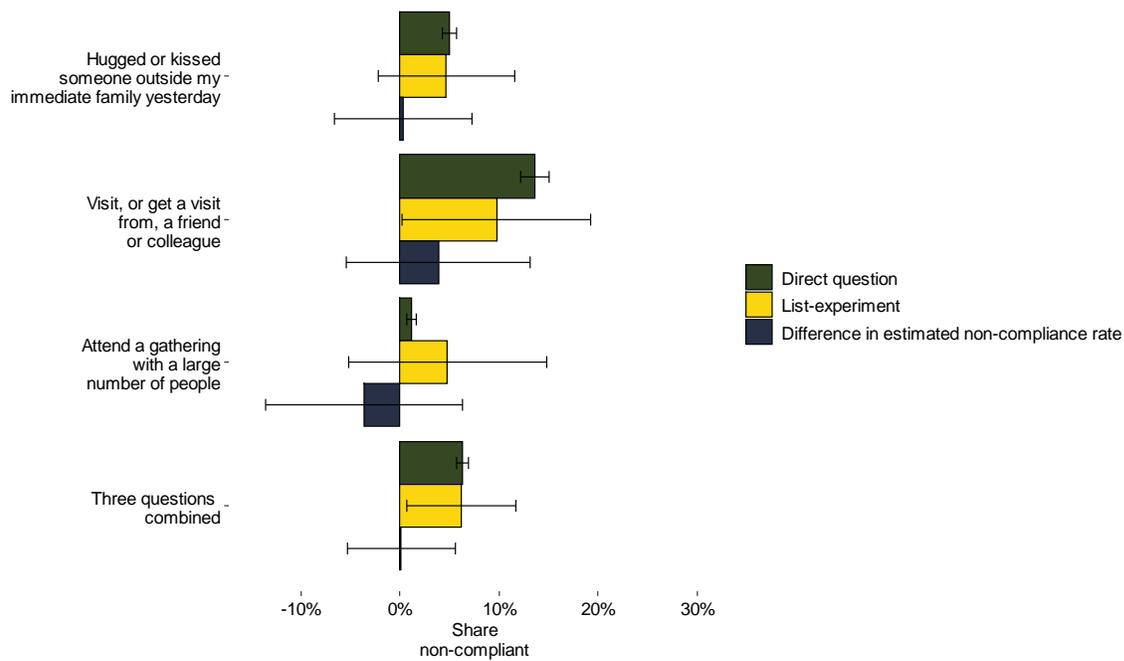
Notes: The error bars indicate the 95 percent confidence intervals. N is 3,497 in the first set of bars and 2,096 in the remaining two middle bars. N is 7,689 in the joint analysis.

Discussion and Conclusion

Our results suggest that social desirability bias does not seem to inflate estimates of compliance with COVID-19 rules and recommendations. This supports suggestions that health authorities may use surveys to track compliance during epidemics including the current and future pandemics (Lau, Yang, Tsui, & Kim, 2003). At the same time, it is worth noting that survey data still has two potential shortcomings. First, although surveys can generate representative estimates more easily than, for example, mobility data from cell phones (which is often limited by app availability, usages and privacy settings), estimates from surveys will still have sampling error. This may in particular be problematic if survey data is used to estimate the reproduction number of COVID-19, where small errors can multiply and lead to biased forecasts given the exponential dynamics of epidemics.[2] Second, we cannot be sure that people are actually behaving as reported in the surveys. Thus, the key conclusion from present studies is that social desirability, specifically as measured though the list-experiment, is not a likely source of bias in the results from surveys during the COVID-19 pandemic. If it was, we would expect to find differences between the list-experimental estimates and the direct questions, as other researchers have found when investigating other types of costly pro-social behaviors (e.g., Comșa & Postelnicu, 2013).

In terms of generalizability, it is important to consider both the setting and the method. In terms of setting, Denmark is a high-compliance context where social and institutional trust is generally very high. It is unclear whether this leads us to over- or underestimate the role played by social desirability bias. On the one hand, strong compliance-norms might mean that very few people would even think of not complying. On the other hand, it might mean that non-compliers are more likely to understand that they are doing something that goes against prevailing norms. In terms of methods, we have relied on online surveys, where social desirability might be less of an issue because you are not communicating directly with another human. As such, it is an open question whether our findings would translate to phone or in-person interviews.

Notes

1. See The Guidelines from the Danish Health Authority. Link (accessed at 2020-05-03): https://www.sst.dk/-/media/Udgivelser/2020/Corona/SST_Corona_A2_Poster-UK-090320.ashx?la=da&hash=234EBC019A46E0ADCE254B6AD0F26A7160FCE5F5.
2. We also test for design effects (for details, see Blair and Imai 2012). The tests give the p-value 1 for the "hugs and kissing someone outside their immediate family", 0.40 for "attend a large social gathering", and 0.88 for "visit a friend or get a visit from a friend", suggesting no clear evidence of design effects.
3. For an example of an analysis that uses survey estimates of compliance to calculate the reproduction number, see "Impact of physical distance measures on transmission in the UK". Link (accessed at 2020-04-03): https://cmmid.github.io/topics/covid19/current-patterns-transmission/comix-impact-of-physical-distance-measures-on-transmission-in-the-UK.html

Acknowledgments

References

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *The Lancet, 395*(10228), 931-934.

Bedsharing and breastfeeding: The academy of breastfeeding medicine protocol #6, Revision 2019" by Blair et al. *Breastfeeding Medicine 2020 15*:7, 478-478

Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis, 20*(1), 47-77.

Comşa, M., & Postelnicu, C. (2013). Measuring social desirability effects on self-reported turnout using the item count technique. *International Journal of Public Opinion Research, 25*(2), 153-172.

Intersection of bias, structural racism, and social determinants with health care inequities Tiffani J. Johnson pediatrics Aug 2020, 146 (2) e2020003657; DOI: 10.1542/peds.2020-003657

Jamieson, K. H., Kahan, D., & Scheufele, D. A. (Eds.). (2017). The Oxford handbook of the science of science communication. Oxford University Press.

Jørgensen, F. J., Bor, A., & Petersen, M. B. (2020). Compliance without fear: Individual-level predictors of protective behavior during the first wave of the COVID-19 pandemic. PsyArXiv, https://psyarxiv.com/uzwgf/

Lau, J. T. F., Yang, X., Tsui, H., & Kim, J. H. (2003). Monitoring community responses to the SARS epidemic in Hong Kong: From day 10 to day 62. *Journal of Epidemiology & Community Health, 57*(11), 864-870.

Pestoff, V., Brandsen, T., & Verschuere, B. (Eds.). (2013). New public governance, the third sector, and co-production. Routledge.

Pfattheicher, S., Nockur, L., Böhm, R., Sassenrath, C., & Petersen, M. B. (2020). The emotional path to action: Empathy promotes physical distancing during the COVID-19 pandemic. https://psyarxiv.com/y2cg5/download?format=pdf

Voorberg, W. H., Bekkers, V. J., & Tummers, L. G. (2015). A systematic review of co-creation and co-production: Embarking on the social innovation journey. *Public Management Review, 17*(9), 1333-1357.

Wimalawansa, S. J. (2020). Global epidemc of coronavirus--COVID-19: What can we do to minimize risks. *European Journal of Biomedical and Pharmaceutical Sciences, 7*(3), 432-438.

Appendix

Online Appendix for: "Do Survey Estimates of the Public's Compliance with COVID-19 Regulations Suffer from Social Desirability Bias?"

The online appendix includes
- Appendix A: Overview of the List Experiments
- Appendix B: Descriptive Statistics
- Appendix C: Response Distribution Tables

Appendix A: Overview of the List Experiments
Table A1 presents an overview of the different list experiments.

Table A1
Overview of the List Experiments

Question	Which of the following activities did you do yesterday?	Which of the following activities are you planning to do in the next few days?	
Sensitive item	*"Hugged or kissed someone outside my immediate family yesterday."*	*"Visit, or get a visit from, a friend or colleague."*	*"Attend a gathering with a large number of people."*
Baseline list	*" I played an online multi-player game with someone outside my immediate family yesterday"*	*" Play a videogame"*	*"Meet my parents"*
	"I did a cross-word puzzle with…	*"Do a cross-word puzzle."*	*"Meet with a sibling."*
	"I watched a movie on a streaming service with..."	*"Watch a movie."*	*"Take a walk with a friend or colleague"*
	"I spoke on the phone with...."	*"Speak with a friend or colleague on the phone."*	*"Talk to a person who is over 90 years old."*
Which survey?	1	2	2
Before/after direct question.	After	Before	Before

Appendix B: Descriptive Statistics

Tables B1 and B2 present descriptive statistics for survey 1 and 2. Respondents are a little older and a little less educated than the population at large.

Table B1

Survey 1. Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
Male	3,515	0.45	0.50	0	1
Age	3,482	48.32	18.32	18	90
North Denmark Region	3,515	0.09	0.28	0	1
Central Denmark Region	3,515	0.24	0.43	0	1
Region of Southern Denmark	3,515	0.20	0.40	0	1
Capital Region of Denmark	3,515	0.34	0.47	0	1
Region Zealand	3,515	0.13	0.33	0	1
Primary school	3,515	0.09	0.28	0	1
General upper secondary education	3,515	0.10	0.30	0	1
Vocational upper secondary education	3,515	0.04	0.20	0	1
Vocational education	3,515	0.23	0.42	0	1
Short-cycle higher education	3,515	0.11	0.31	0	1
Medium-cycle higher education	3,515	0.19	0.39	0	1
Bachelor's degree	3,515	0.10	0.29	0	1
Long cycle higher education	3,515	0.14	0.35	0	1
Employed in the private sector	3,515	0.28	0.45	0	1
Employed in the public sector	3,515	0.19	0.40	0	1
Self-employed	3,515	0.04	0.20	0	1
Enrolled in education	3,515	0.13	0.33	0	1
Unemployed	3,515	0.05	0.22	0	1
Pensioner/early retirement	3,515	0.27	0.44	0	1
Other	3,515	0.04	0.20	0	1

Table B2

Survey 2. Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
Male	2,069	0.47	0.50	0	1
Age	2,096	49.02	16.52	18	97
North Denmark Region	2,096	0.11	0.31	0	1
Central Denmark Region	2,096	0.24	0.43	0	1
Region of Southern Denmark	2,096	0.21	0.40	0	1
Capital Region of Denmark	2,096	0.28	0.45	0	1
Region Zealand	2,096	0.16	0.37	0	1
Primary school	2,096	0.10	0.30	0	1
Upper secondary education	2,096	0.10	0.30	0	1
Vocational education	2,096	0.24	0.43	0	1
Short-cycle higher education	2,096	0.13	0.33	0	1
Medium-cycle higher education	2,096	0.28	0.45	0	1
Long cycle higher education	2,096	0.15	0.36	0	1
Other education	2,096	0.002	0.04	0	1
Employed	1,640	0.55	0.50	0	1
Self-employed	1,640	0.05	0.21	0	1
Enrolled in education	1,640	0.07	0.25	0	1
Long-term sick leave	1,640	0.02	0.14	0	1
Unemployed	1,640	0.08	0.28	0	1
Pensioner/early retirement	1,640	0.23	0.42	0	1
Other	1,640	0.01	0.08	0	1

Appendix C: Response distribution tables
Tables C1-C3 present response distribution tables for the three different list experiments.

Table C1
Hugged or kissed someone outside my immediate family yesterday.

Response Value	Control group		Treatment group	
	Frequency	Proportion (%)	Frequency	Proportion (%)
0	377	22%	372	21%
1	721	41%	741	42%
2	448	26%	430	24%
3	138	8%	151	9%
4	59	3%	46	3%
5			32	2%
Total	1,743		1,772	

Table C2
Visit, or get a visit from, a friend or colleague.

Response Value	Control group		Treatment group	
	Frequency	Proportion (%)	Frequency	Proportion (%)
0	50	5%	52	5%
1	160	15%	170	16%
2	406	39%	355	34%
3	283	27%	300	29%
4	145	14%	133	13%
5			42	4%
Total	1,044		1,052	

Table C3
Attend a gathering with a large number of people

Response Value	Control group		Treatment group	
	Frequency	Proportion (%)	Frequency	Proportion (%)
0	430	41%	454	43%
1	324	31%	304	29%
2	184	18%	156	15%
3	74	7%	60	6%
4	39	4%	37	4%
5			34	3%
Total	1,051		1,045	