

It depends on how you ask: measuring bias in population surveys of compliance with COVID-19 public health guidance

Shane Timmons ^{1,2} Frances McGinnity,^{1,3,4} Cameron Belton,¹ Martina Barjaková,¹ Peter Lunn^{1,5}

► Supplemental material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2020-215256>).

¹Economic and Social Research Institute, Dublin, Ireland

²School of Psychology, Trinity College Dublin, Dublin, Ireland

³Trinity College Dublin, Dublin, Ireland

⁴School of Sociology, Trinity College Dublin, Ireland

⁵Department of Economics, Trinity College Dublin, Ireland

Correspondence to

Shane Timmons, Economic and Social Research Institute, Whitaker Square, Sir John Rogerson's Quay, Dublin, Ireland; shane.timmons@esri.ie

Received 1 August 2020

Revised 30 September 2020

Accepted 5 October 2020

ABSTRACT

Objective Accurate measurement of compliance with COVID-19 guidance is important for public health policy and communications. Responses to surveys, however, are susceptible to psychological biases, including framing effects and social desirability. Our aim was to measure the effects of these biases on estimates of compliance with public health guidance (eg, hand-washing, social distancing).

Design We conducted two online experiments (n=1800) and varied whether questions were framed positively or negatively (eg, 'I always wash my hands. . .' vs 'I don't always wash my hands. . .'). We also varied the degree to which anonymity was assured, via a 'list' experiment.

Results Reported compliance, despite being generally high, was reduced by negatively framing questions and increasing anonymity using a list experiment technique. Effect sizes were large: compliance estimates diminished by up to 17% points and 10% points, respectively.

Conclusion Estimates of compliance with COVID-19 guidance vary substantially with how the question is asked. Standard tracking surveys tend to pose questions in ways that lead to higher estimates than alternative approaches. Experimental tests of these surveys offer public health officials greater insight into the range of likely compliance estimates to better inform policy and communications.

INTRODUCTION

Containing the spread of COVID-19 requires widespread compliance with public health guidance, including hand hygiene and distancing from others.¹ These behaviours are hard to measure objectively, so governments and public health officials rely on estimates from tracking surveys. We present two experiments showing that these estimates depend strongly on how questions are asked. The experiments were commissioned by Ireland's Department of Health, to support the Behavioural Change Subgroup of the National Public Health Emergency Team (NPHE).

The way questions are asked can affect responses in surveys.^{2,3} Potential sources of variation include order effects, where the order in which questions or response alternatives are presented influences respondents' answers⁴ and survey format effects (ie, how the survey is administered).⁵ These 'method effects', whereby some variation in outcome is associated with how it is measured, are problematic if responses are systematically biased.⁶ Here, we investigated two specific potential sources of bias: question framing⁷ and social desirability.⁸

We varied whether survey questions were framed positively or negatively. Logically, if a survey asks people whether they regularly wash their hands and 90% say 'yes', the same survey should find that 10% report *not* washing their hands. However, positive or negative framing can alter responses.^{9,10}

Social desirability refers to the tendency for survey respondents to over-endorse items that they perceive others judge favourably.¹¹ If participants believe that COVID-19 risk mitigation behaviours are socially desirable, some who do not follow guidance may be reluctant to respond truthfully. Thus, reported compliance in surveys may be inflated.

METHOD

To measure framing effects, we randomised survey respondents to answer positively or negatively framed questions about the same behaviour. To estimate social desirability bias, we used a 'list experiment'.¹² In this method, a first group of randomly assigned participants views a list of items, comprised of non-target items and one target item. Participants are not asked *which* items apply to them, only *how many*.¹³ A second (control) group views only the non-target items and is asked the same question. Thus, the difference between the average response to both lists indicates the proportion of participants who endorsed the target item. The method confers anonymity: researchers infer the prevalence of the target behaviour without individuals endorsing it explicitly. By subsequently asking the control group directly about the target behaviour, prevalence under anonymity in the experiment can be compared to prevalence measured via a direct question.¹⁴

The study was conducted in line with institutional ethics policy.

Participants

We recruited 1800 adults from an online panel held by a leading market research company. Socio-demographic characteristics approximated census figures well, as summarised in online supplemental material. Timmons *et al* provide details on how recruitment from this panel compares to a probability sample.¹⁵ Eight hundred completed the first experiment in mid-June. The remainder completed the second experiment 2 weeks later. A national tracking survey showed no change in the target behaviours over this period. Participants undertook the experiments as part of a 20-min online study programmed using Gorilla Experiment Builder.¹⁶ They were paid €6.



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Timmons S, McGinnity F, Belton C, *et al*. *J Epidemiol Community Health* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jech-2020-215256

Materials, design and procedure

In each experiment, participants were randomly assigned to one of two conditions. In the 'list' condition ($n_1=402$, $n_2=502$), participants viewed a list containing the target behaviour (eg, hand-washing) along with three non-target behaviours. They reported how many items applied to them. For example,

- ▶ I have been keeping in touch with friends and family via the internet or by phone.
- ▶ I am watching less TV (or streaming TV shows less) than usual.
- ▶ I have a household pet that I have been spending more time with (eg, taking the dog for a walk more often).
- ▶ I wash my hands with soap and water for a full 20 seconds (or I use hand sanitiser) when I return home from being out or touch a surface other people might have touched.

Item order was randomised. The online supplemental material provides further details and full materials. In the 'direct' condition ($n_1=398$, $n_2=498$), participants viewed the list of three non-target items and reported how many applied to them. A direct question about the target item followed, for example,

...does the below action apply to you?

- ▶ I wash my hands with soap and water for a full 20 seconds (or I use hand sanitiser) when I return home from being out or touch a surface other people might have touched.

In the first experiment, items were framed positively, as above. In the second experiment, the same items were framed negatively (eg, 'I don't wash my hands. . .' (sic)). Each experiment included three target items and hence three sets of questions, presented in random order: hand-washing, distancing and meeting others (experiment 1); hand-washing, distancing and mask-wearing (experiment 2). These items reflected contemporaneous public health guidance. As framing was tested only for hand-washing and distancing, we focus on these items for the purpose of this paper. Results of the list experiment for meeting others and mask-wearing are reported in online supplemental material.

RESULTS

We preregistered directional hypotheses for the effects of social desirability and non-directional hypotheses for framing. The preregistration, data and analysis code are available at <https://osf.io/3ukqg/>. The online supplemental material contains additional details on the analysis, including robustness checks.

Hand-washing

When asked the direct, positively framed question, 91% of participants reported following hand-washing guidance, which matched contemporaneous national tracking data.¹⁷ Figure 1 compares conditions. When the question was framed positively, reported compliance was lower in the list condition than the direct condition, $t(451.9)=1.78$, $p=0.038$, $d=0.13$. When framed negatively, there was no difference, $t(612.5)=0.09$, $p=0.464$, $d=0.01$. However, a lower proportion of participants reported washing their hands when asked the direct, negatively framed question compared to the direct, positive question, $t(891.4)=3.67$, $p<0.001$, $d=0.24$. There was no difference between frames in the list condition, $t(900)=0.21$, $p=0.834$, $d=0.01$.

Distancing

Ninety-two percent reported keeping 2 m from others when directly asked the positively framed question, again matching the national survey. There was no significant reduction in the list condition, $t(466.9)=0.83$, $p=0.204$, $d=0.06$ (figure 1). However, when the question was framed negatively, there was evidence of lower reported compliance in the list condition, $t(600.6)=1.39$,

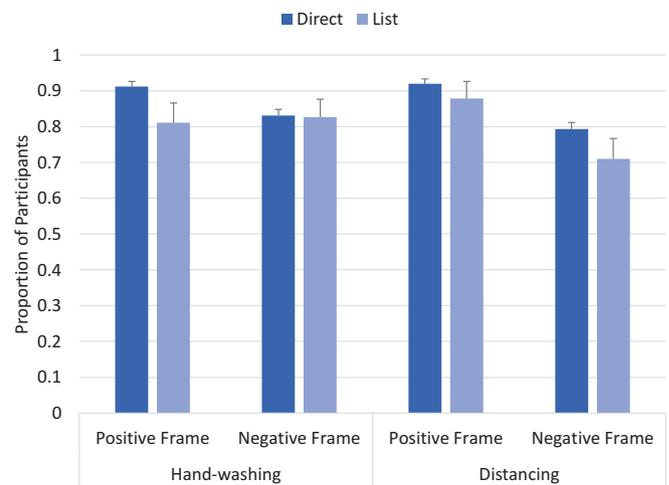


Figure 1 Proportion of participants endorsing target items. Error bars are the SE. Error bars for the 'list' conditions are larger due to the combined variances when calculating the proportion of indirect endorsements.

$p=0.082$, $d=0.09$. In general, negative framing reduced reported compliance in the direct, $t(869.5)=5.56$, $p<0.001$, $d=0.37$, and list conditions, $t(896.7)=2.27$, $p=0.023$, $d=0.15$.

Socio-demographic differences

Because list experiments are analysed using difference-in-means estimators (as above), standard models that incorporate socio-demographic variables as individual-level covariates are not possible.¹⁴ Instead, we repeated the above analysis for subgroups by gender, age, education¹⁸ and residential area (urban/rural). Note that these subgroup tests have reduced statistical power. Three exploratory comparisons were statistically significant (table 1); all other comparisons were non-significant (details at <https://osf.io/3ukqg/>). These results suggest that effects of social desirability and frame depend not only on the relevant behaviour (eg, hand-washing and social distancing), but also on respondents' socio-demographic characteristics (such as gender, age and whether they live in an urban or rural area).

Table 1 Significant differences by socio-demographic subgroups

Health behaviour	Socio-demographic characteristic	Finding
Hand-washing	Gender	More men reported complying with hand-washing advice when asked directly than in the list condition, but only when the question was framed positively (89% vs 76%), $t_{\text{pos}}(241.1)=1.75$, $p=0.040$, $d=0.17$.
Distancing	Age	Among younger participants (aged under 50 years old), there was a difference between the direct and list conditions when framed both positively (88% vs 74%), $t(231.5)=1.61$, $p=0.054$, $d=0.16$, and negatively (75% vs 64%), $t(344.2)=1.40$, $p=0.082$, $d=0.12$.
	Living area	Rural participants displayed significant social desirability bias when the distancing question was framed positively (93% vs 81%), $t(174.1)=1.64$, $p=0.051$, $d=0.19$, and negatively (83% vs 63%), $t(211.5)=2.27$, $p=0.012$, $d=0.24$.

DISCUSSION

Tracking surveys indicate high compliance with COVID-19 public health guidance in Ireland. We found equivalent estimates when we posed positively framed questions directly to participants. However, attempts to reduce social desirability bias decreased reported compliance by up to 10% points. Varying question framing produced differences of up to 17% points, with negative frames generating lower estimates. These effects were large: roughly doubling and more than doubling measured non-compliance, respectively. Thus, estimates of compliance depend strongly on how the question is asked. Notably, the effects varied across target behaviours. For example, whereas social desirability did not affect reported distancing in the positive frame,¹⁹ it did in the negative frame. Hand-washing showed the opposite pattern: social desirability bias affected hand-washing in the positive frame, but not in the negative one.

Our experiments do not show which estimates most accurately reflect behaviour. However, since list experiments counter social desirability bias, the results suggest that direct questions that measure self-reported compliance probably overestimate true compliance. Why reported compliance is lower when questions are framed negatively is unclear. Multiple psychological mechanisms could be advanced and future research may determine which frame is most accurate.

Controlling the spread of COVID-19 will depend on continued engagement with public health advice. We have shown that compliance may appear artificially high if surveys employ direct, positively framed questions, as tracking surveys typically do. Researchers might improve the quality of evidence from compliance surveys by asking multiple forms of questions, permitting triangulation of more accurate estimates, as recommended in research on method effects.⁶ Experimental methods can further help to reveal the potential scale of inaccuracy.²⁰ Controlled testing of survey questions can help public health officials and communications teams to identify behaviours that require stronger promotional messaging.

What is already known on this subject

- ▶ Compliance with public health guidance is vital for containing the spread of COVID-19 but is difficult to measure objectively, meaning public health officials rely on national tracking surveys.
- ▶ People are sensitive to how questions are framed and sometimes overstate their agreement with survey items if they think others judge those items favourably; these biases risk inflating tracking survey estimates.

What this study adds

- ▶ We show that negatively framed survey questions (eg, 'I don't always keep 2 metres from others in public...') more than double non-compliance estimates compared to more standard, positively framed questions (eg, 'I always keep 2 metres from others...').
- ▶ Conferring survey respondents greater anonymity doubles estimates of non-compliance with public health advice compared to standard tracking surveys.
- ▶ Experimental tests of survey questions offer a way for public health officials to better understand rates of non-compliance with COVID-19 guidance.

Acknowledgements We thank Deirdre Robertson for help in designing items for experiment 1 and Ciarán Lavin for help in creating negative frames in experiment 2. We also thank Mathew Creighton for guidance and the Behavioural Change Subgroup of the NPHE for helpful comments.

Contributors ST developed the study concept, designed the materials in collaboration with FMG, performed the data analysis and interpretation, and drafted the manuscript. MB assisted with material design and reviewing relevant literature. CB programmed the experiment. CB, FMG and PL provided critical revisions. All authors approved the final version of the manuscript for submission.

Funding Both experiments were funded through a research programme commissioned by Ireland's Department of Health, in support of the Behavioural Change Subgroup of the National Public Health Emergency Team (NPHE).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the COVID-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

ORCID iD

Shane Timmons <http://orcid.org/0000-0002-0200-5927>

REFERENCES

- 1 Anderson RM, Heesterbeek H, Klinkenberg D, *et al.*, How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 2020;395:931–4.
- 2 Schaeffer NC, Dykema J, Advances in the science of asking questions. *Ann Rev Sociol* 2020;46:37–60.
- 3 Choi BC, Pak AW. Peer reviewed: a catalog of biases in questionnaires. *Prev Chronic Dis* 2005;2:1. Available <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1323316/>
- 4 Strack F. 'Order effects' in survey research: activation and information functions of preceding questions. In: Schwarz N, Sudman S (eds.). *Context effects in social and psychological research*. New York, NY:Springer, 1992: 23–34.
- 5 Schwarz N, Strack F, Hippler HJ, *et al.*, The impact of administration mode on response effects in survey measurement. *Appl Cogn Psychol* 1991;5:193–212.
- 6 Maul A, Method effects and the meaning of measurement. *Front Psychol* 2013;4:169.
- 7 Chong D, Druckman JN, Framing theory. *Annu Rev Polit Sci* 2007;10:103–26.
- 8 Krumpal I, Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quant* 2013;47:2025–47.
- 9 Kahneman D, Tversky A. Rational choice and the framing of decisions. *Journal of Business* 1986;59:251–78.
- 10 Druckman JN, The implications of framing effects for citizen competence. *Political Behav* 2001;23:225–56.
- 11 Furnham A, Response bias, social desirability and dissimulation. *Pers Individ Dif* 1986;7:385–400.
- 12 Creighton MJ, Jamal A, Does Islam play a role in anti-immigrant sentiment? An experimental approach. *Soc Sci Res* 2015;53:89–103.
- 13 Glynn AN, What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opin Q* 2013;77:159–72.
- 14 Blair G, Imai K. Statistical analysis of list experiments. *Political Anal* 2012;20:47–77.
- 15 Timmons S, Barjaková M, Robertson D, *et al.* Public understanding and perceptions of the COVID-19 test-and-trace system. *Econ Social Res Inst (ESRI) Res Ser* 2020 Aug;
- 16 Anwyll-Irvine AL, Massonnié J, Flitton A, *et al.*, Gorilla in our midst: an online behavioral experiment builder. *Behav Res Methods* 2020;52:388–407.
- 17 Department of health, government of Ireland. *Gov.ie* [Internet]. updated 14 July 2020. Available <https://www.gov.ie/en/collection/6b4401-view-the-amarach-public-opini-survey/> (accessed 17 Jul 2020)
- 18 McGinnity F, Creighton M, Fahey É. Hidden versus revealed attitudes: a list experiment on support for minorities in Ireland. *Econ Social Res Inst (ESRI) and Irish Human Rights and Equality Commission (IHREC)* 2020.
- 19 Larsen MV, Petersen MB, Nyrup J. Do survey estimates of the public's compliance with COVID-19 regulations suffer from social desirability bias? *PsyArxiv* 2020.
- 20 Lunn PD, Belton CA, Lavin C, *et al.*, Using behavioral science to help fight the coronavirus: a rapid, narrative review. *J Behav Public Administration* 2020;3:1.