

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Protocol for creating new warnings on cigarette packs and evaluating their efficacy in a randomised experimental setting
AUTHORS	BEN LAKHDAR, Christian; Deplancke, Antoine; LE LEC, Fabrice; Massin, Sophie; Piermatteo, Anthony; Vaillant, Nicolas

VERSION 1 – REVIEW

REVIEWER	Gallopel-Morvan, Karine EHESP School of Public Health, Rennes
REVIEW RETURNED	11-Feb-2020

GENERAL COMMENTS	<p>This article deals with the effect of new warnings on smokers and nonsmokers and with a novel way of evaluate them, through monetary incentives. More precisely, this paper presents the protocol of a research that will be conducted in order:</p> <ul style="list-style-type: none">. to test innovative warnings which content (pictorial and textual) is based on nudge and behavioural theories plus a “warning” which aims is to motivate smokers to vape (harm reduction message);. to develop a new method to evaluate the effect of warnings on behaviours (in regard with strong limitations of self-reported methods).. to test the effect of warnings when they are displayed on tobacco branded vs. plain packaging. <p>This paper is very relevant as it presents a very original protocol method to evaluate warnings. To my opinion, the main added-value of this protocol paper is the new method of evaluation that is proposed by authors. Regarding the innovativeness of warnings, it’s difficult to know as messages are not shown nor described in details.</p> <p>I would suggest some revisions / comments that could increase the quality of this paper before publishing it in BMJ open:</p> <p>Major comments</p> <ul style="list-style-type: none">. This paper presents the research protocol of a study that will be conducted in 2020. This aim (presenting a protocol research) should be clearly mentioned in the title, in the introduction and in the summary of the “strengths and limitations” of the paper.. Could authors justify why the costs of a quitting method book and the donation to an NGO may predict smoking behaviours (it’s quite clear for attitude toward smoking but not for behaviours)?. Beyond the effectiveness of warnings, different variables may influence the money that people will give to an NGO (donation) and will pay for a book: budget, prices of these kinds of books in shops, notoriety of the NGO for instance. Will you check that? How?
-------------------------	---

	<p>. On page 4, readers of this paper can have the feeling that non-declarative methods have never been conducted to evaluate the impact of tobacco warnings, which is not true. Indeed, eye-tracking, fMRI and other methods have been used before to evaluate warnings (see Maynard, O. M., McClernon, F. J., Oliver, J. A., & Munafò, M. R. (2019). Using neuroscience to inform tobacco control policy. <i>Nicotine and Tobacco Research</i>, 21(6), 739-746; Green, A. E., Mays, D., Falk, E. B., Vallone, D., Gallagher, N., Richardson, A., ... & Niaura, R. S. (2016). Young adult smokers' neural response to graphic cigarette warning labels. <i>Addictive behaviors reports</i>, 3, 28-32.; Droulers, O., Gallopel-Morvan, K., Lacoste-Badie, S., & Lajante, M. (2017). The influence of threatening visual warnings on tobacco packaging: Measuring the impact of threat level, image size, and type of pack through psychophysiological and self-report methods. <i>PloS one</i>, 12(9)). Then authors of this paper should explain the added-value of the non-declarative method they suggest, compared to previous ones that have already been used to evaluate warnings.</p> <p>. More details should be given regarding methods:</p> <ul style="list-style-type: none"> - The tested stimuli could be shown on photos (the 40 messages could be presented in a supplementary file) - I suggest adding table(s) that will present the experimental design of the research (samples per group, number of smokers / non smokers per group, type and number of warnings per group, etc.) - Is it a “real” donation and a “real” quitting method book that people will give /buy (It’s unclear if the donation and the book will be given/bought after the experimentation or if people will just have to declare the amount of money?). If it is a declaration, behavioural intentions and not real behaviours will then be measured. - Could you describe the “typical questionnaire measures” that you will use in your research? <p>. One of the aim of the paper is to test new warnings. I suggest explaining your contribution regarding novelty compared to J. Hoek’s research: Hoek, J., Hoek-Sims, A., & Gendall, P. (2013). A qualitative exploration of young adult smokers’ responses to novel tobacco warnings. <i>BMC public health</i>, 13(1), 609.</p> <p>Minor comments:</p> <ul style="list-style-type: none"> . I suggest changing the title of the paper (presentation of a protocol research etc.) and keywords that are too large (I suggest for instance: tobacco, warnings, evaluation, nudge, incentives) . To better justify why you selected a message that recommends an harm reduction method to quit, you could have a look on these papers: Thrasher, J. F., Islam, F., Davis, R. E., Popova, L., Lambert, V., Cho, Y. J., ... & Hammond, D. (2018). Testing cessation messages for cigarette package inserts: Findings from a best/worst discrete choice experiment. <i>International journal of environmental research and public health</i>, 15(2), 282; Moodie, C. (2018). Adult smokers' perceptions of cigarette pack inserts promoting cessation: a focus group study. <i>Tobacco control</i>, 27(1), 72-77) . Authors speak of “health” warnings whereas these messages are not always health-oriented (including those that will be tested in this research). Then I suggest deleting the word “health” before warnings. . I think there is a mistake in the abstract “[...] plain pack without warning, plain pack with existing warning, and pack with existing
--	---

	<p>warning". As authors will test new warnings, I think that's not "existing" that should be written in this sentence.</p> <p>. It is said that "perceived effectiveness" will be evaluated: but effectiveness on what (behaviours, attitudes, etc.)?</p> <p>Different words are used for the same idea (I think): effectiveness, efficacy, efficiency. I suggest using only one word to convey the same idea.</p>
--	--

REVIEWER	Eric Crosbie University of Nevada, Reno
REVIEW RETURNED	25-Feb-2020

GENERAL COMMENTS	<p>Overall this is a nicely written paper that is attempting to create a new system to evaluate the efficacy of cigarette pack health warning labels and plain packaging. The authors incorporate the use a nudge approach to design new warnings and look to test these in 2020. There are questions related to France as a case study, the selection of experts, the choice of variables, etc but without the analysis it is unclear what the results will actually show. I wonder if this type of research is publishable without a results section and instead if this may be better suited to publish once the study is performed? However, if the BMJ accepts exploratory research and methodology papers I think the authors just need to address my minor comments outlined below. Thus, only minor revision if journal accepts this type of research.</p> <p>Introduction:</p> <p>Page 3, lines 17-18: It is not that plain packaging is the main communication and advertising tool it is more that it is one of the last forms of advertising given that several countries are banning most parts of tobacco advertising and furthermore the package has been labeled a "mobile billboard" so it has important characteristics as well. I would suggest the authors characterize it to a degree in this way.</p> <p>Page 3, line 30: The authors contend that one could wonder about the types of warnings that could be more effective but some of the literature has examined this issue. For example, some research illustrates that the more graphic/repulsive the image the better in terms deterring people from smoking or helping current smokers quit. Please see various work by Jim Thrasher, David Hammond and Geoffrey Fong.</p> <p>Methods:</p> <p>Page 4, lines 9-12: This is true but there are instances when plain packaging is done alone in isolation and not attached with other pieces of legislation (e.g. Ireland, Canada, Uruguay, etc). Why not study these case studies to minimize or eliminate this explanation?</p> <p>Page 4, lines 12-14: The authors are correct to point out the difference in effect between smokers and non-smokers but what about within smokers those that are heavy smokers vs light/causal smokers?</p> <p>Page 6, line 7: Why select 5 messages per approach, why not 3 or other numbers? Basically what is the rationale?</p>
-------------------------	--

	<p>Page 6, line 15: Who exactly are these French experts and how are they chosen to rank efficacy? Why not just refer to the literature to support selections?</p> <p>Page 9, line 21: What about a question concerning smoker types (e.g. heavy vs light smokers)?</p> <p>Discussion:</p> <p>Page 10: I just wonder about country specifics of France and how this variable is incorporated in the model? To check reliability and see how this approach will work in other contexts the authors may want to look deeper into other factors that may influence these results. For example, if the smoking prevalence in France is twice as much as Australia and U.S. like authors said wouldn't something even like smoking prevalence rates matter? There may be other factors as well.</p> <p>Minor Edits:</p> <p>Page 3, line 6: Suggest changing to "the most preventable cause of death worldwide"</p> <p>Page 7, line 51: Change studies to study in sentence "this type of studies"</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: karine Gallopel-Morvan

Institution and Country:

EHESP School of public health

France

Please state any competing interests or state 'None declared': non declared

Please leave your comments for the authors below

This article deals with the effect of new warnings on smokers and nonsmokers and with a novel way of evaluate them, through monetary incentives. More precisely, this paper presents the protocol of a research that will be conducted in order:

- . to test innovative warnings which content (pictorial and textual) is based on nudge and behavioural theories plus a "warning" which aims is to motivate smokers to vape (harm reduction message);
- . to develop a new method to evaluate the effect of warnings on behaviours (in regard with strong limitations of self-reported methods).
- . to test the effect of warnings when they are displayed on tobacco branded vs. plain packaging.

This paper is very relevant as it presents a very original protocol method to evaluate warnings. To my opinion, the main added-value of this protocol paper is the new method of evaluation that is proposed by authors. Regarding the innovativeness of warnings, it's difficult to know as messages are not shown nor described in details.

I would suggest some revisions / comments that could increase the quality of this paper before publishing it in BMJ open:

On behalf of the research team, we would like to warmly thank Pr. Gallopel-Morvan for these comments and suggestions. We are indebted to her better positioning our research protocol within the existing literature. We hope that our answers will be interesting and satisfactory to her.

Major comments

. This paper presents the research protocol of a study that will be conducted in 2020. This aim (presenting a protocol research) should be clearly mentioned in the title, in the introduction and in the summary of the “strengths and limitations” of the paper.

You are absolutely right. We now specify that this paper is a research protocol directly in the title, and also make reference to that fact in the introduction as well as in the “strengths and limitations” section.

. Could authors justify why the costs of a quitting method book and the donation to an NGO may predict smoking behaviours (it's quite clear for attitude toward smoking but not for behaviours)?

You are right in pointing out this issue. We were not precise enough in the original manuscript when relating the incentivized measures with attitudes, intentions, and behaviour. We now state the following in the revised manuscript (p.7):

“Note that, strictly speaking, the WTP for the quitting book does not measure smoking (or quitting) behaviour per se. Measuring actual behaviours would require to follow subjects for a long-term range, which would be very challenging in practice. Nevertheless, our indicator does measure the strength, immediately after exposure to the warnings, of the subject's intention to quit without being affected by the hypothetical biases that may characterize declarative answers.”

We believe that “buying a book” is still a form of behaviour (because this is incentivized and some subjects may get the book for real instead of money – see below for the clarification on that point), but indeed what is being measured is an intention to quit in the near future rather than an actual quitting action.

Regarding the distinction between behaviour and intentions, we agree that, for smokers at least, buying the book (or more exactly the willingness to pay for the book) clearly reveals more their intention to quit than an actual smoking cessation behaviour. Nevertheless, smoking cessation can be apprehended as a process that may include a series of micro-actions within a long sequence (decision, implementation of this decision, resisting temptations...), and as a result, it is difficult to consider it to be a locally isolated behaviour (see DiClemente et al., 1991; Prochaska & DiClemente, 1983 or more recently Ham, 2007). In this perspective, buying the book could be understood here to be the first step in the action of quitting, as would buying pharmaceutical products for nicotine substitution or seeking medical help, etc. Moreover, buying the book is, figuratively and literally, more costly than expressing the intention to quit smoking on a questionnaire, it may be more predictive of a subsequent quitting behavior than said intention according to commitment theory (Kiesler, 1971).

Smoking cessation is in all probability a course of action that starts earlier than the actual behaviour and, to some extent, lasts for as long as the ex-smoker does not smoke. Some evidence of this is given by de Vries et al. (2013). This means that it would be extremely difficult to measure the whole behavioural chain that goes from some warnings to definite cessation. In a way, the intermediate behaviour of buying a book or not seems to us a convenient proxy. We have now defined with more precision in the manuscript how the monetary value ascribed to the book relates to intentions to quit and behaviour.

In a broader sense, our concern is that non-incentivized methods have an increased risk of bias because of the declarative nature of the measure. Incentivizing a decision related to smoking behaviour is assumed to reduce that bias. More specifically, our concern with traditional methods is that (most of the time although not always) the measurement of attitudes, intentions, etc. is entirely declarative. This may result in a hypothetical bias that is perfectly documented in the scientific literature. We refer to that explicitly in the manuscript. Moreover, we believe that one particular aspect of the hypothetical bias, the so-called social desirability bias, may be very strong in the case of tobacco smoking given the strong social pressure/norms that surround it these days. In particular, we fear that instead of measuring the idiosyncratic response to the warnings of individuals, we may measure an evolution of the social desirability of non-smoking. The channel would be as follows: some types of warnings (but not all of them) reinforce the social desirability of not smoking, or of quitting, (as experienced by subjects) which leads subjects to declare being more willing to quit, etc., even though they only do so because it would be more subjectively costly to state the opposite since it would go against social expectations. It is this line of reasoning that we tried to condense into one sentence, where references 25 and 26 were mentioned (p.4). We have now expanded this sentence with what we believe are clarifications:

“The third one is a common problem in behavioural science: the answers to a questionnaire may have a hypothetical (25) or a social desirability bias (a tendency to conform to what is perceived as the experimentalist’s expectations) (26): the problem may be particularly acute in our case as warnings may affect the perceived social desirability of quitting for instance without genuinely modifying the individual’s intention in this regard.”

A second source of error is the fact that when the self-interest of individuals is at stake they tend to make more cognitive effort in answering questions or making decisions. And this generally comes

with a reduction of noise in the data. See for instance Camerer and Hogarth (1999).

A third element, sometimes called the Wallis and Friedman (1942, p. 177-183) criticism (see also Roth 1993 on the topic), is that individuals may not know exactly how they would react if the situation were not hypothetical. They may believe that they would react strongly, but in fact would not. Behavioural science literature provides ample evidence of this partial inability of people to predict their own behaviour, people are often too optimistic about their behaviour (see for instance Armor & Taylor, 1998; Dunning, 2007 for the general case, and Lipkus & Shepperd, 2009 for smoking behaviour). We believe that our incentivized measure will help address these concerns.

Overall, we believe that setting-up, in addition to the traditional questions about intentions, attitudes, etc., an incentivized measure of the intention to quit for smokers, and attitude towards tobacco smoking for non-smoker is a relevant way to test the robustness of hypothetical measures. We do not think nor claim that the incentivized measure completely annihilates those biases but we think that it at least reduces them substantially. Moreover, it is of interest (and that is part of our contribution) to determine experimentally the level of correlations between both.

We have added some of the sentences mentioned above to clarify the importance of incentive compatibility to shed light on this issue.

References:

- Armor, D. A., & Taylor, S. E. (1998). Situated optimism: Specific outcome expectancies and self-regulation. *Advances in Experimental Social Psychology*, 30, 309–379.
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19(1-3), 7-42.
- de Vries, H., Eggers, S. M., & Bolman, C. (2013). The role of action planning and plan enactment for smoking cessation. *BMC Public Health*, 13(1), 393.
- DiClemente, C. C., Prochaska, J. O., Fairhurst, S. K., Velicer, W. F., Velasquez, M. M., & Rossi, J. S. (1991). The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages of change. *Journal of consulting and clinical psychology*, 59(2), 295-304.
- Dunning, D. (2007). Prediction: The inside view. In A. W. Kruglanski and E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (Second edition, pp. 69-90). New York: Guilford Press.

Ham, O. K. (2007). Stages and processes of smoking cessation among adolescents. *Western Journal of Nursing Research*, 29(3), 301-315.

Kiesler CA. The psychology of commitment: experiments linking behavior to belief. New York: Academic Press; 1971. 190 p. (Social psychology).

Lipkus, I., & Shepperd, J. (2009). College smokers' estimates of their probabilities of remaining a smoker in the near future. *Journal of Health Psychology*, 14(4), 547–555.

Poon, C. S. K., D. J. Koehler, and R. Buehler. "On the psychology of self-prediction: Consideration of." *Judgment and Decision Making*, 9.3 (2014): 207-225.

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of consulting and clinical psychology*, 51(3), 390-395.

Roth, A.E. "On the Early History of Experimental Economics," *Journal of the History of Economic Thought*, 15, Fall, 1993, 184-209.

Wallis, W. A., & Friedman, M. (1942). The empirical derivation of indifference functions. *Studies in mathematical economics and econometrics in memory of Henry Schultz*, 175-189.

. Beyond the effectiveness of warnings, different variables may influence the money that people will give to an NGO (donation) and will pay for a book: budget, prices of these kinds of books in shops, notoriety of the NGO for instance. Will you check that? How?

Absolutely. Many variables will influence the monetary value people ascribe to the book (or the donation to the NGO), most of them with no relation to smoking attitudes, intentions, etc. Yet, the great strength of experimental methods is to allow to control for that: as formally shown by Rubin (1979) – who “simply” formalized experimental methodologies in many evidence-based disciplines –, the randomized allocation of subjects into treatments guarantees that the subsamples of every treatment are statistically similar (on average, because of the law of large numbers). As a result of randomization, the probability of these variables playing a role in the average behaviour in one treatment (compared to the other) is low (that is actually what p-values measure in theory).

Nevertheless, because there is always the possibility of some sampling variations, we intend to control for some of these variables in our analysis by using typical econometric “ceteris paribus” estimations. This is mostly true for the sociodemographic variables.

For the specific variables you mentioned, we will reduce their effect as much as possible on several grounds: regarding the notoriety of the NGO, all non-smokers will be provided detailed information about the NGO, so that, in the end, we expect all subjects to be similarly familiar with the NGO. The NGO is “*la fondation du souffle*” (the foundation for breath).

Regarding the price of the books in shops, it so happens that France has a price control on books (authorized variations are limited to 5%), so there is an “official price” for this book. The WTP measure is calibrated to stay under this price (so that no one – for instance knowing the price of the book – could think that it would be better to get the money and buy the book themselves). The book is: Allen Carr (2011). *La méthode simple pour en finir avec la cigarette*. Pocket editions. (8.90€).

But ultimately, despite the care taken in experimentally mitigating the role of these variables or econometrically accounting for the sociodemographic characteristics of the subjects, the most relevant method is the randomization of subjects to treatments that guarantees, on average, that this perturbation will be similar in all treatments.

Reference:

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74(366a), 318-328.

. On page 4, readers of this paper can have the feeling that non-declarative methods have never been conducted to evaluate the impact of tobacco warnings, which is not true. Indeed, eye-tracking, fMRI and other methods have been used before to evaluate warnings (see Maynard, O. M., McClernon, F. J., Oliver, J. A., & Munafò, M. R. (2019). Using neuroscience to inform tobacco control policy. *Nicotine and Tobacco Research*, 21(6), 739-746; Green, A. E., Mays, D., Falk, E. B., Vallone, D., Gallagher, N., Richardson, A., ... & Niaura, R. S. (2016). Young adult smokers' neural response to graphic cigarette warning labels. *Addictive behaviors reports*, 3, 28-32.; Droulers, O., Gallopel-Morvan, K., Lacoste-Badie, S., & Lajante, M. (2017). The influence of threatening visual warnings on tobacco packaging: Measuring the impact of threat level, image size, and type of pack through psychophysiological and self-report methods. *PLoS one*, 12(9)). Then authors of this paper should explain the added-value of the non-declarative method they suggest, compared to previous ones that have already been used to evaluate warnings.

You are absolutely right, we oversimplified the issue. We have modified the manuscript accordingly (see page 4), including references to the neuroscience study on the topic you suggested as well as explaining the difference more clearly with our concerns. The core of the matter, we believe, is that most, if not all, of the neurophysiological studies on this topic use declarative measures (attitudes, intentions) as a benchmark, to relate these declarative measures to neurophysiological signals. These studies are definitely interesting as we can learn much about the psychological/neurological processes that generates this or that reaction to a given warning, but they do not necessarily provide a more robust measure of the psychological reaction. For instance, it is plausible that extremely shocking graphical warnings have a strong emotional impact, which generates (for some areas) a peak in physiological activity. The fact that warnings are “violent” may also suggest to subjects that the expected social behaviour is not to smoke and that others are intensely committed to that (hence

the strong warnings), leading to an increase in the social desirability bias. But this strong emotional impact may not generate a change in (real) behaviour. In other words, our incentivized measure (the WTP task) may be more appropriate to quantify the projection of individuals regarding their future behaviour than the immediate response to the warning's stimuli as is the case in standard and neurophysiological studies.

More generally, we believe that our incentivized measure, survey questions, and neurophysiological responses are complementary to a large extent. As highlighted by Green et al. (2019), neuroscience techniques complement and extend traditional methods (observational experiments, surveys, and questionnaires). Our research protocol and above all our behavioural response measurement – the willingness to pay – is entirely in line with this conclusion but goes a bit further in the sense that it is a real act, namely a purchase, and is observed not neurophysiologically but behaviourally. As explained in the protocol, the participant (smoker or not) who will sacrifice part of the amount of money given to her/him in the experiment will bear real consequences from her/his behaviour.

Another dimension in which our protocol is complementary to neurophysiological measures is sample size. Neurophysiological measures are generally quite costly, and as a result, sample sizes are rather small. In contrast, the size of our sample will allow us to get statistically sound results.

. More details should be given regarding methods:

- The tested stimuli could be shown on photos (the 40 messages could be presented in a supplementary file)

Absolutely, we now present the messages in a supplementary file.

- I suggest adding table(s) that will present the experimental design of the research (samples per group, number of smokers / non smokers per group, type and number of warnings per group, etc.)

We have added a table in the paper summarizing the experimental design (Table 2 p.9).

- Is it a “real” donation and a “real” quitting method book that people will give /buy (It's unclear if the donation and the book will be given/bought after the experimentation or if people will just have to declare the amount of money?). If it is a declaration, behavioural intentions and not real behaviours will then be measured.

Indeed, subjects, depending on their choices in the experiment, will receive some money and/or a book (or alternatively they will receive money or make a donation an NGO). This is a key element of the design: their choices will have real consequences. And it is in that sense that we believe – see above – that our design measures actual behaviour.

In practice, during the iterative Multiple Price List (iMPL) tasks, subjects will respond to binary choices as follows:

Question number n	Option A: Receiving the book	Option B: Receiving a given amount of money	Your choice: A or B?
-------------------	---------------------------------	--	-------------------------

If this question is randomly selected for implementation at the end of the experiment, then the answer the subject has chosen for this particular question will be given to her/him. For instance, if s/he chose the book (option A), s/he will receive the book (plus the show up fee). On the contrary, if s/he chose Option B, s/he will receive a given amount of money in addition to the show up fee.

As the question used for implementation is randomly determined (and does not depend on what people do), it is in their best interest to respond truthfully to all questions as all of these have an equal chance of being selected.

For non-smokers, the questions will look as follows:

Question number n	Option A: 8.90 € are donated to the NGO	Option B: You receive a given amount of money	Your choice: A or B?
-------------------	--	--	-------------------------

Same as for smokers, a question will be randomly selected. If at this particular question the subject has chosen A then s/he will receive no additional money (just the show up fee), but 8.90€ will be sent to the NGO (and this will indeed be the case). On the contrary, if s/he chose B, then s/he will receive a given amount of money in addition to the show up fee.

- Could you describe the “typical questionnaire measures” that you will use in your research?

These questionnaire measures include consensual and standard questions. As we specified in our contribution, we rely on three different types of existing studies described below.

Questions about smoking are taken from the Health Barometer of Public Health France. The Health Barometer is a large representative health survey questioning the French on their health behaviour. The questionnaire is available here:

We will thus be able to see whether there is any difference between our sample and the general population.

In addition, we use the very standard Fagerström test for cigarette dependence with two questions. (Svicher A, Cosci F, Giannini M, Pistelli F, Fagerström K. Item Response Theory analysis of Fagerström Test for Cigarette Dependence. *Addict Behav.* 2018 Feb;77:38–46.)

Concerning time and risk preferences, we use the questions and results from this paper: Falk, Armin and Becker, Anke and Dohmen, Thomas and Huffman, David and Sunde, Uwe, The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. IZA Discussion Paper No. 9674. Available at SSRN: <https://ssrn.com/>

Concerning attention, recall, intention and perceived effectiveness, we use the Francis et al (2017) systematic review of measures used in pictorial cigarette pack warning experiments. *Nicotine and Tobacco Research*, 19(10), 1127-1137. Using a review and meta-analysis leads us to focus on the most consensual (since they are frequent) questions used in the literature.

Subsection "Collected variables" has been rewritten accordingly (p.9-10).

. One of the aim of the paper is to test new warnings. I suggest explaining your contribution regarding novelty compared to J. Hoek's research: Hoek, J., Hoek-Sims, A., & Gendall, P. (2013). A qualitative exploration of young adult smokers' responses to novel tobacco warnings. *BMC public health*, 13(1), 609.

Thank you for this reference. We have added it in our manuscript (reference 46, citation p.6). In this paper, new warnings are built and appeal to smokers' rationality by providing information about the objective dangers of smoking (novel and immediate health risks, social risks, and tobacco as a toxin). Their efficacy is tested by in-depth interviews with 17 young adult smokers.

Our protocol differs in two dimensions: the type of warnings tested and the method used to test them. Regarding the first point, in our protocol, we build new warnings that appeal to the "system 1" of individuals, through nudges. In other words, we test different types of interventions than those usually used, namely rationality and awe/emotional shock. The presence, following your suggestion, of a supplementary file showing the warnings (text and pictures) will help clarify this point.

Regarding the second point, we test the warnings not with in-depth interviews, but through traditional surveys and some, more innovative, incentive compatible measures, allowing us, among other things, to rely on a more important sample.

Minor comments:

. I suggest changing the title of the paper (presentation of a protocol research etc.) and keywords that are too large (I suggest for instance: tobacco, warnings, evaluation, nudge, incentives)

We have added "protocol" in the title. Unfortunately, we are not free to choose the keywords. They are provided by BMJ in a closed list. We tried to choose them as best we could.

. To better justify why you selected a message that recommends an harm reduction method to quit, you could have a look on these papers: Thrasher, J. F., Islam, F., Davis, R. E., Popova, L., Lambert, V., Cho, Y. J., ... & Hammond, D. (2018). Testing cessation messages for cigarette package inserts: Findings from a best/worst discrete choice experiment. *International journal of environmental research and public health*, 15(2), 282; Moodie, C. (2018). Adult smokers' perceptions of cigarette pack inserts promoting cessation: a focus group study. *Tobacco control*, 27(1), 72-77)

Thank you for these references, which indeed consider NRT as a means of managing the craving for nicotine. We go further by considering this same use of nicotine with ENDS as a way to quit smoking. We have added these articles in our references (references 43 and 44; citation p.6).

. Authors speak of "health" warnings whereas these messages are not always health-oriented (including those that will be tested in this research). Then I suggest deleting the word "health" before warnings.

You are totally right. Thank you for this suggestion. We have deleted the word "health" when inappropriate.

. I think there is a mistake in the abstract "[...] plain pack without warning, plain pack with existing warning, and pack with existing warning". As authors will test new warnings, I think that's not "existing" that should be written in this sentence.

The wording used in the original manuscript was indeed confusing. We have replaced "existing" by "conventional", by which we mean regulatory warnings used in the past in France or abroad (Belgium). These warnings are used as controls to test the efficacy of the new (innovative) warnings that we have created. So, finally, there are three control groups (see p.7):

- 1- Plain packs with no warning
- 2- Plain packs with conventional regulatory warnings
- 3- Branded packs with conventional regulatory warnings

. It is said that “perceived effectiveness” will be evaluated: but effectiveness on what (behaviours, attitudes, etc.)?

In fact, we will measure the perceived effectiveness as defined in Francis et al. (2017), which give the following items (which are about behaviours):

- For all:
 - Perception of warning’s message motivational value for others: not starting smoking
 - Perception of warning’s message motivational value for others: cutting down on smoking
 - Perception of warning’s message motivational value for others: quitting smoking
- For smokers:
 - Perception of warning’s message motivational value for myself: quitting smoking
- For non-smokers:
 - Perception of warning’s message motivational value for myself, if I was a smoker: quitting smoking

We have changed the phrasing to “Perceived effectiveness (not starting smoking, cutting down on or quitting smoking)” to be more precise (p.10).

Different words are used for the same idea (I think): effectiveness, efficacy, efficiency. I suggest using only one word to convey the same idea.

Thank you for this suggestion. We have used “efficacy” as much as possible knowing that "efficacy" refers to a theoretical ideal and that "effectiveness" is its translation in the real world (see Amin et al., (2004). The difference between effectiveness and efficacy of antimalarial drugs in Kenya. *Tropical Medicine and International Health*, 9(9), 967-974.)

We have also emphasized the great efficiency of warnings in the sense that they are inexpensive for public finances and certainly achieve their objective at the lowest cost, or even at zero cost.

In a nutshell, the accepted uses of the words are (and we have tried to be as precise as possible in our phrasing in the manuscript):

- Efficacy: how theoretically (and in ideal conditions such as an experiment) strong the impact is
- Effectiveness: how strong the impact is in real (in the field) conditions
- Efficiency: how strong the impact is compared to the cost of implementation

We have not included a footnote or a sentence in the manuscript about this terminology but would be happy to do so if you believe (or the editor believes) that it is necessary for the understanding of the

typical reader of the journal.

Reviewer: 2

Reviewer Name: Eric Crosbie

Institution and Country: University of Nevada, Reno

Please state any competing interests or state 'None declared': None

The research team would like to thank Professor Crosbie for these comments and suggestions. We hope our responses will satisfy him.

Please leave your comments for the authors below

Overall this is a nicely written paper that is attempting to create a new system to evaluate the efficacy of cigarette pack health warning labels and plain packaging. The authors incorporate the use a nudge approach to design new warnings and look to test these in 2020. There are questions related to France as a case study, the selection of experts, the choice of variables, etc but without the analysis it is unclear what the results will actually show. I wonder if this type of research is publishable without a results section and instead if this may be better suited to publish once the study is performed? However, if the BMJ accepts exploratory research and methodology papers I think the authors just need to address my minor comments outlined below. Thus, only minor revision if journal accepts this type of research.

Introduction:

Page 3, lines 17-18: It is not that plain packaging is the main communication and advertising tool it is more that it is one of the last forms of advertising given that several countries are banning most parts of tobacco advertising and furthermore the package has been labeled a "mobile billboard" so it has important characteristics as well. I would suggest the authors characterize it to a degree in this way.

Indeed, we fully agree. More than an active policy aiming at reducing smoking behaviour, it is a way to prevent the last opportunity for marketing for the tobacco industry. Moreover (and this is our main concern in this study as we focus on the efficacy of warnings), it may interact with the warnings. In this sense, comparing commercial packs with warnings and plain packs with warnings would be more about the negative effect of commercial packs than the positive effect of plain packs (who, as you pointed out, seek to mitigate marketing actions from the tobacco industry rather than actively deter people from smoking).

It is also true that the commercial packs could work as an "attention grabber" and prevent people to

pay (enough) attention to the warnings. Said differently, warnings are perhaps more visible on the plain pack than on the commercial pack and that could play a role.

We have modified the sentence following your suggestion (p.3).

Page 3, line 30: The authors contend that one could wonder about the types of warnings that could be more effective but some of the literature has examined this issue. For example, some research illustrates that the more graphic/repulsive the image the better in terms deterring people from smoking or helping current smokers quit. Please see various work by Jim Thrasher, David Hammond and Geoffrey Fong.

That's definitely true, but what we try to do here is to propose (and test, since, to be honest, as "empiricists" we do not claim they will be better) new *types of warnings*, that is warnings that do not appeal to the individual's rationality ("smoking kills" for instance) or emotional response (awe obtained by "graphically explicit" warnings for instance). By new types of warnings, we refer to other types of motivational leverages (not rational decision or emotional reaction) as behavioural science has demonstrated the existence of a large spectrum of types of motivations.

We have clarified this point in the main text and added some appropriate references (page 3, ref 12-14).

Methods:

Page 4, lines 9-12: This is true but there are instances when plain packaging is done alone in isolation and not attached with other pieces of legislation (e.g. Ireland, Canada, Uruguay, etc). Why not study these case studies to minimize or eliminate this explanation?

That could be done, for instance with proper statistical or econometric models (for instance differences between countries that adopt plain packaging at a specific point). There is one major drawback though: the first one is that the proper identification of the effect of plain packaging alone is extremely difficult to obtain (it may take some time for people to get rid of past marketing actions and subjective associations between a brand and positive feelings for instance), and most of the time many other things (including tobacco related legislation) change at the same time. It seems to us, that at the macro-level (countries), it is difficult to draw conclusions from a statistical analysis of the general level of tobacco consumption.

Regarding our specific study, our main aim is to test those new warnings. The plain/commercial pack is more an additional robustness treatment (for the controls) than the core of our study.

Page 4, lines 12-14: The authors are correct to point out the difference in effect between smokers and non-smokers but what about within smokers those that are heavy smokers vs light/causal smokers?

We fully agree that there may be some heterogeneous effects of the warnings, especially given the type of smokers. We intend to test this through several channels:

- First, we ask smokers about the intensity of their tobacco use (typically, for cigarette smokers, the number they smoke per day) as well as estimate their level of addiction through the Fagerström test for cigarette dependence with two questions. (Svicher A, Cosci F, Giannini M, Pistelli F, Fagerström K. Item Response Theory analysis of Fagerström Test for Cigarette Dependence. *Addict Behav.* 2018 Feb;77:38–46.)
- We will be able to compare subsamples of smokers in different treatments to see if the effect is conditional on being of a given type (light/heavy). This can be done by limiting the sample used for comparisons to heavy smokers for instance.
- We will use econometric regression models that will provide control variables (and interaction variables of the type (Type of Smoker X Treatment) that will allow us to know whether the average effect is disproportionately related to a type of smoker.

Page 6, line 7: Why select 5 messages per approach, why not 3 or other numbers? Basically what is the rationale?

There are several reasons to focus on 5 messages. The first is that standard practice in these studies is to expose subjects to between 4 and 7. We chose the “mean” as we are interested in the type of warnings (rather than in individual warnings) and we did not want to go beyond 5 to stay within the standard methods to be able to compare our results with previous studies. Choosing more would also have meant expecting a reduction in attention for our participants (attention fatigue). And finally, as we had to design the messages and fund the production of graphics and images, budget size also came into play.

We are now more specific about that point in the manuscript (p.6):

“Each of them will have to choose between two pictures for each text message. That will provide us with 5 pairs of message and image for each type. Probably in order to maintain participants’ attention and for budgetary reasons, experimental studies in this field generally use 4 to 7 warnings (43,44,46), so our study is comparable in this respect. Indeed, as we intend to test types of warnings and the relevance of motivational leverages rather than a specific warning, we chose to use a mean of the number of warnings used in a typical study.”

Page 6, line 15: Who exactly are these French experts and how are they chosen to rank efficacy? Why not just refer to the literature to support selections?

We cannot refer to the literature to determine which type of “nudges” (empowerment, engagement or cognitive dissonance) is better than the others because, to our knowledge, there are no studies that directly test their relative strengths... We believe this to be true for the general field and we are almost

certain of it regarding tobacco warnings.

The experts are individuals who are officially recognized as such by public institutions. They are members of:

- *Santé Publique France* (Public Health France):
<https://www.santepubliquefrance.fr/a-propos/notre-organisation/une-organisation-au-service-des-programmes/conseil-scientifique>
- Observatoire Français des Drogues et Toxicomanies (French Monitoring Center for Drugs and Drug Addiction):
<https://www.ofdt.fr/ofdt/college-scientifique/>

They are all officially involved in public health matters and tobacco smoking control policies.

We will ask this panel of French public health experts to choose warnings among the 40 that we will submit to them, via a dedicated website. They will have to choose, for the same text message, which image they believe is most likely to bolster the non-smoker in his choice and encourage the smoker to modify his behaviour. See the supplementary file (requested by the other referee) with our 40 initial warnings.

Page 9, line 21: What about a question concerning smoker types (e.g. heavy vs light smokers)?

Yes, we have planned to fill in this kind of information. We will use the questions of a large French health survey on this subject (the Health Barometer). We have made this more clear (p.9-10).

Discussion:

Page 10: I just wonder about country specifics of France and how this variable is incorporated in the model? To check reliability and see how this approach will work in other contexts the authors may want to look deeper into other factors that may influence these results. For example, if the smoking prevalence in France is twice as much as Australia and U.S. like authors said wouldn't something even like smoking prevalence rates matter? There may be other factors as well.

You are definitely right. We cannot claim that our results will necessarily be valid for other countries or cultural areas. Yet, despite this limitation, we would like to point out two things. First, if our experiment was to show that some other type of warning than the current regulatory one is substantially better – even if it is only the case for France –, that would suggest that it is possible to find more appropriate psychological leverage, and would encourage similar studies to be run in all countries (a cost benefit

analysis is unambiguously in favour of generalizing this approach if shown to be relevant in one area). Second, as with any experiment, we cannot presume its external validity, but we can at least assume that in culturally similar areas (European French-speaking areas, or Western Europe, or Latin Europe, etc.), it is likely that similar effects are at play (even though they remain to be tested).

Regarding the effect of other (local) variables, you are absolutely right: the simple fact that more than twice as many French people smoke (in relative terms) than Australians or Americans implies that these populations of smokers do not have the same composition. Nevertheless, and we agree that it applies to our study, any locally run experimental study should always be taken with this limitation in mind and be replicated in a different context to be able to generalize the results.

Minor Edits:

Page 3, line 6: Suggest changing to “the most preventable cause of death worldwide”

Page 7, line 51: Change studies to study in sentence “this type of studies”

Thank you. Changes done.

VERSION 2 – REVIEW

REVIEWER	Gallopel-Morvan, Karine EHESP School of Public Health, Rennes
REVIEW RETURNED	07-May-2020

GENERAL COMMENTS	<p>Thank you for the version 2 of this paper which has been improved. It is now clear that it is the presentation of a protocol, an in addition it is an innovative and an interesting protocol.</p> <p>Please find below the minor changes I suggest for the version 3 of the paper:</p> <p>. The contributions of the paper could be clearer in the introduction (page 3): I think your paper adds new ideas on how to evaluate public health policies (first main contribution). Indeed, public health evaluation is very complex (see Smith, R. D., & Petticrew, M. (2010). Public health evaluation in the twenty-first century: time to see the wood as well as the trees. <i>Journal of Public Health</i>, 32(1), 2-7.), so new and innovative methods are welcome, including those that go beyond limits of classical declarative methods (2nd main contribution).</p> <p>. Page 3: some more information could be given on warnings that are currently displayed in France (and in Europe): size, location, etc.</p>
-------------------------	---

	<p>. Page 6: if I well understood, warnings have been already created by a professional designer (and the study conducted on experts too). Then the appropriate tense is the past and not the future.</p> <p>. Page 6: as warnings used in Belgium and France are the same (they come from the European database), other warnings could be presented in order to avoid previous exposure (warnings of countries outside the EU, https://tobaccolabels.ca/).</p> <p>. Page 10: not only “economists tend to dismiss declarative data and rely almost exclusively on “revealed” preferences”. Some researchers in neuroscience and in psychology also think the same.</p> <p>. As readers of the BMJ open journal are mainly English-spoken people, I suggest to change the reference n°13 (paper in French) with this one in English (on the same topic): Gallopel-Morvan K., Gabriel P., Le Gall-Ely M., Rieunier S. et Urien B. (2011), The use of visual warnings in social marketing: the case of tobacco, Journal of Business Research, 64 (1), 7-11</p> <p>Good luck for the version 3 of the paper!</p>
--	---

REVIEWER	Eric Crosbie University of Nevada, Reno
REVIEW RETURNED	24-Apr-2020

GENERAL COMMENTS	We thank the authors for answering questions and incorporating feedback.
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Thank you for the version 2 of this paper which has been improved. It is now clear that it is the presentation of a protocol, an in addition it is an innovative and an interesting protocol.

Please find below the minor changes I suggest for the version 3 of the paper:

. The contributions of the paper could be clearer in the introduction (page 3): I think your paper adds new ideas on how to evaluate public health policies (first main contribution). Indeed, public health evaluation is very complex (see Smith, R. D., & Petticrew, M. (2010). Public health evaluation in the twenty-first century: time to see the wood as well as the trees. Journal of Public Health, 32(1), 2-7.), so new and innovative methods are welcome, including those that go beyond limits of classical declarative methods (2nd main contribution).

Thank you for this advice and for this reference that we have added in the text.

. Page 3: some more information could be given on warnings that are currently displayed in France (and in Europe): size, location, etc.

We now give more elements of the displaying of warning but above all we specify from which European directive these standards come. Thank you, this is an important point.

. Page 6: if I well understood, warnings have been already created by a professional designer (and the study conducted on experts too). Then the appropriate tense is the past and not the future.

Yes absolutely, thank you.

. Page 6: as warnings used in Belgium and France are the same (they come from the European database), other warnings could be presented in order to avoid previous exposure (warnings of countries outside the EU, <https://tobaccolabels.ca/>).

Thank you for this suggestion. We believe it is important to use warnings used in countries that are not too culturally different from France to avoid the risk of biasing the results. As the subject pool is mainly composed of students or former students and the median age is presumably around 20, the risk of previous exposure is certainly minimized taking “old” warnings.

. Page 10: not only “economists tend to dismiss declarative data and rely almost exclusively on “revealed” preferences”. Some researchers in neuroscience and in psychology also think the same.

Yes absolutely. We added these other research fields in, the sentence.

. As readers of the BMJ open journal are mainly English-spoken people, I suggest to change the reference n°13 (paper in French) with this one in English (on the same topic): Gallopel-Morvan K., Gabriel P., Le Gall-Ely M., Rieunier S. et Urien B. (2011), The use of visual warnings in social marketing: the case of tobacco, Journal of Business Research, 64 (1), 7-11

We have changed the reference, thank you.

Good luck for the version 3 of the paper !

Thank you very much for all your comments!

VERSION 3 – REVIEW

REVIEWER	Gallopel-Morvan, Karine EHESP School of Public Health, Rennes
REVIEW RETURNED	18-May-2020
GENERAL COMMENTS	The version 3 of this paper has been improved. I suggest to publish this protocol reserach.