HOW TO CATCH A CHEATER

ANSWER
AB1543.986

RELEVANCE & RESEARCH QUESTION:

Online surveys are self-administered by respondents seeking to receive incentives for completing questionnaires. Some respondents use minimal cognitive effort in order to quickly complete the survey in return for incentives. However, this can trigger behaviour such as not reading the questions carefully, racing through the survey or intentionally cheating, resulting in poor data quality. This paper aims to investigate the behaviour of cheaters among online respondents from a non-probability-based panel analyzing seven techniques for detecting cheaters, applied in different ways in order to find an efficient methodology that leads to the elimination of the greatest possible number of cheaters without removing honest panelists.

Methods & Data:

We used data from 2 web surveys conducted in Italy (during January 2019) on members of our own panel, Opinioni.net, which is composed of 20,558 active panelists. The 2 surveys considered in our study have the following characteristics: a sample size of 1,073 for the first dataset and of 1,004 for the second, the same population target and a food consumption topic. Sample members were stratified by geographic area, gender and age in order to be representative of the Italian population.
In both questionnaires, we asked a particular question: "Are you affected or have you been affected in the past by one or more of the following long-term illnesses or pathological conditions?". We considered suffering of "Allergies" the target variable of our studies.
The first survey was used as training set to determine a method for identifying cheaters. In particular, we analyzed the estimates of the target variable in each check and in any combination thereof, in comparison with the estimation of the same question from the Multi-purpose Survey on Families: "Health conditions and appeal to health services" conducted by Istat in 2016 in order to assess the quality of our data. Once the method was defined, we validated it using the second survey as test set.

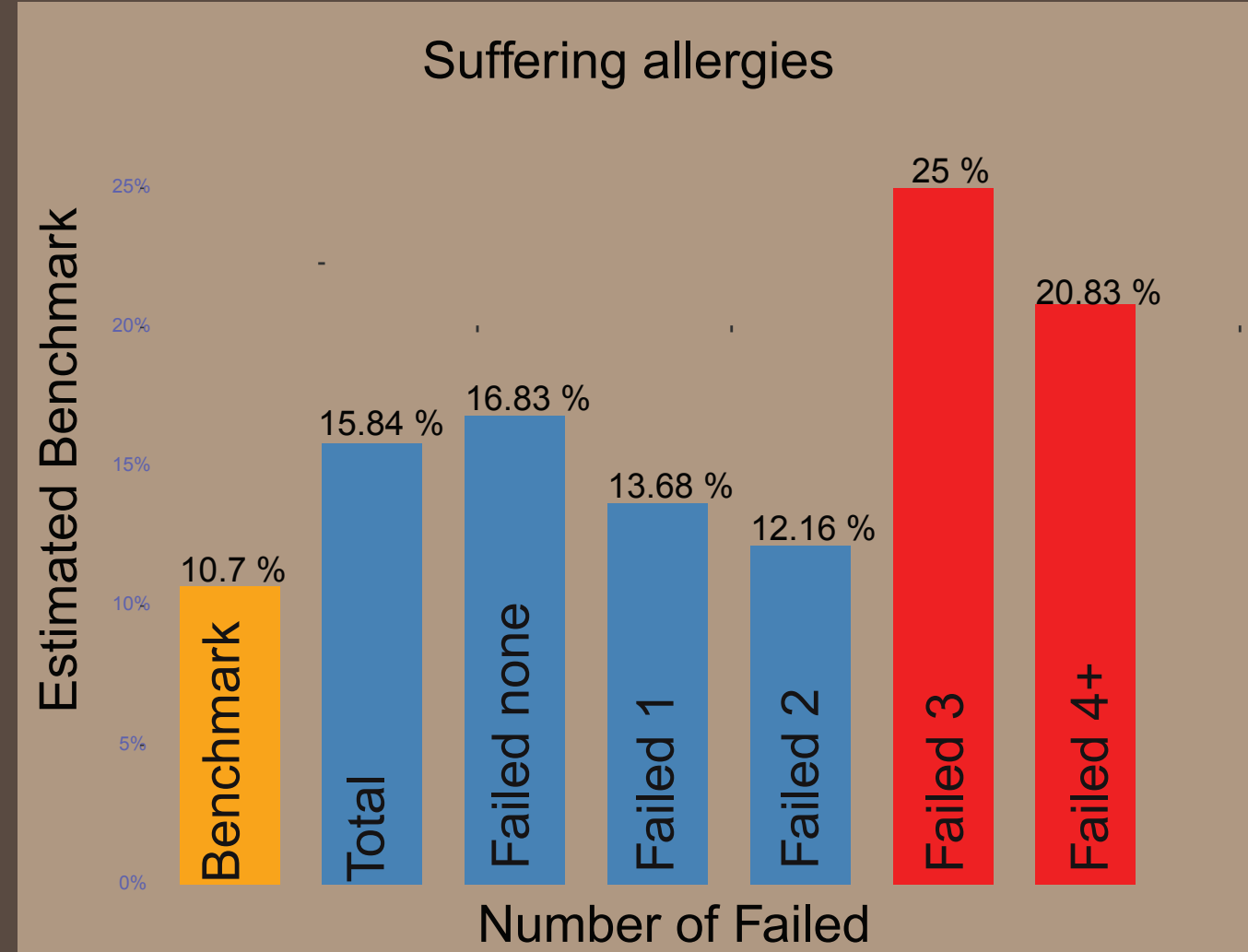## THE TECHNIQUES USED TO DETECT CHEATERS ARE:

1 **Direct instruction in question body** (direct_1, direct_2): these are questions with a direct instruction in the body which aims to check if the text of the application has been carefully read. In this case we used two Radio Buttons Question Type:
1. direct_1: "To continue click on Friday..."
2. direct_2: "To demonstrate that you have read this instruction, please do not answer the question below. Instead click on the Next button to continue filling out the questionnaire".

2 **Unlikely events** (unlik): At the beginning of a questionnaire there are often some screening questions aimed at detecting if the respondent has the characteristics required to access the survey.
The respondent may state that he has all the required characteristics, even if this is not the case, with the aim of proceeding with the questionnaire and gaining the final incentive. To identify this type of cheater it is sufficient to insert, among the screening questions, a question asking whether the respondent has one or more characteristics, which he would not normally have.

3 **Trap questions (Fake brand/Names)** (fake): These consist of incorporating fictitious (ghost) brands or names in a question.
For one survey we chose a Radio Buttons Question Type: "Have you ever heard of NAME OF FAKE SERVICE?" (Yes/No)
For the other we chose to incorporate a fake brand name to other real ones in a yes/no grid.

4 **Bad Open Questions:** (open): In a mandatory open-ended question, participants can use an inappropriate or unreliable answer (eg, 'Asdfhjkl') as a way to indicate the lack of a meaningful answer or as a means of simply not leaving the space blank.

5 **Consistency checks** (coer_1, coer_2): The validation checks consist of two or more questions related to each other that are placed at different points of the questionnaire. In these circumstances the answer to the second question should correspond or at least not contradict the first one.

6 **Speeder checks** (time): Speeders are survey participants who finish too quickly. The problem with this kind of method is to set the cut-off to define "how quick is too quick".
We considered a percentage of mean/median time of compilation:
1. 33% of the mean time of compilation after excluding outliers.
2. 48% of the median time of compilation.
For each path in the survey we set the cut-off as the mean of these two values.

7 **Straightlining checks** (straigh_1, straigh_2, straigh_3): Straightlining occurs when survey respondents give identical (or with predictable pattern) answers to items in a battery of questions using the same response scale.
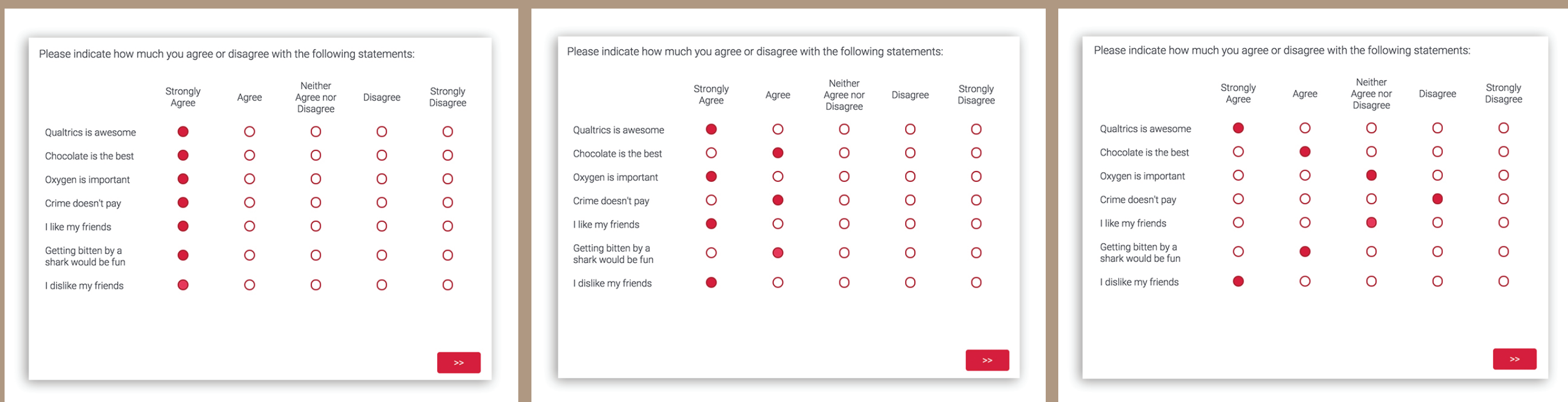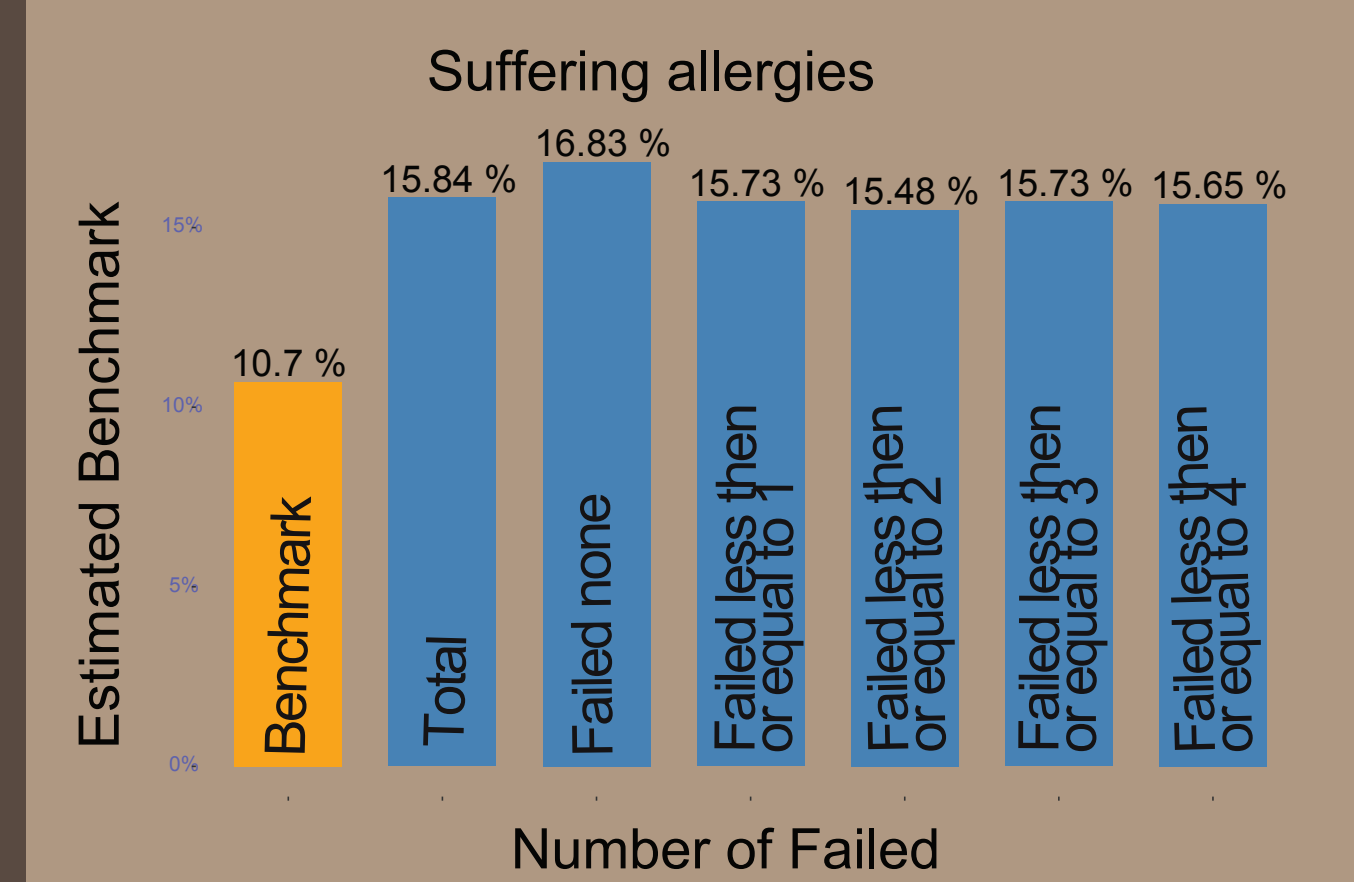
To catch this type of cheater, we evaluated the average of the differences in absolute value between adjacent scores combined with the time taken to complete the set of questions. When the answers are presented in a straight line (for example 1,1,1,1...), the average of the differences should be around 0. Instead, when they have a "zig-zag" shape, like 1,2,1,2... or 1,2,3,2,1..., this score should be around 1. This kind of behaviour reduces the time of completion because the respondent is not engaged in the questionnaire. In trials, it was found that the average CAWI responder needs 300 milliseconds to understand a single word in a sentence. This factor multiplied by the number of words in the question determines the time needed to read and understand the question properly. For these reasons, those respondents for whom the average of the differences was 0 or 1 and the time of compilation was less than the required reading time, were considered to have failed the straightlining checks.

Partecipants Failing each question
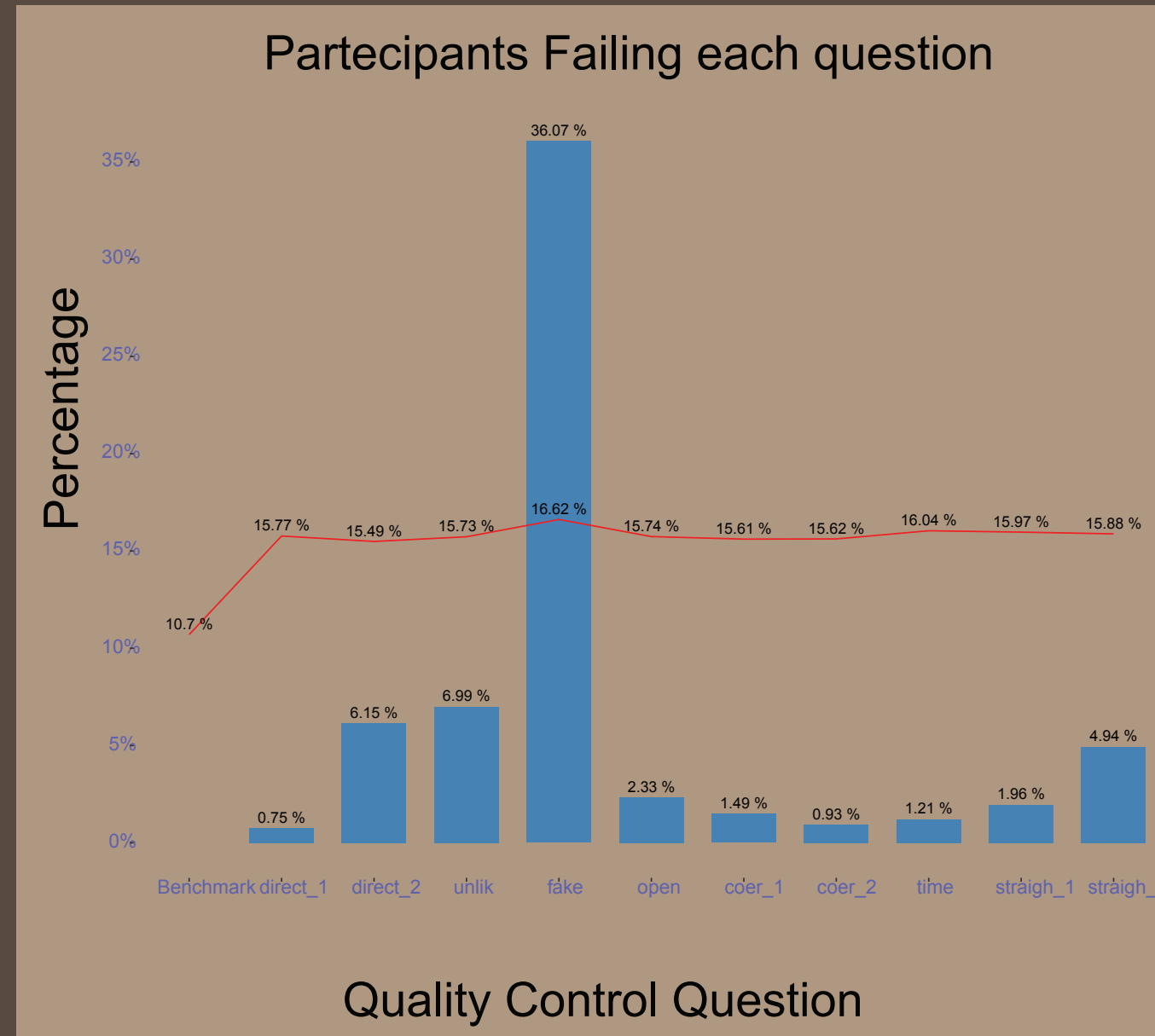
This histogram represents the percentage of respondents who fail each control, while the red line shows the estimates of the target variable for the sample with the exception of those who fail the single check. There is a high percentage of those who fail the trap questions (fake brand/names), which also corresponds to a higher estimate of the target variable than that of other checks. This type of question could be tricky because there will always be some percentage of respondents who think they've heard of a brand even when the brand doesn't exist, but they're honestly mistaken. Punishing innocence could be a means of biasing the sample and in fact there is a rise in the estimate of the target variable. Another interesting result is the difference in the percentage of respondents who fail the direct instruction in the body of the question: for the second check the answer options were "Yes", "No" or "I do not know", while for the first check the answer options were the day of the week which were de-standardized compared to the classic response options and this may have caught the respondent's attention.

Partecipants Failing each question

In this case we chose to incorporate a fake brand to other real ones in a yes/no grid and with respect to the previous data set, there is a considerable reduction in the number of those who fail the check. We checked for straightlinings in 3 batteries placed at the beginning, in the middle and at the end of the questionnaire with the result that there is an increase in the number of those who fail the control based on the longer time spent completing the questionnaire.
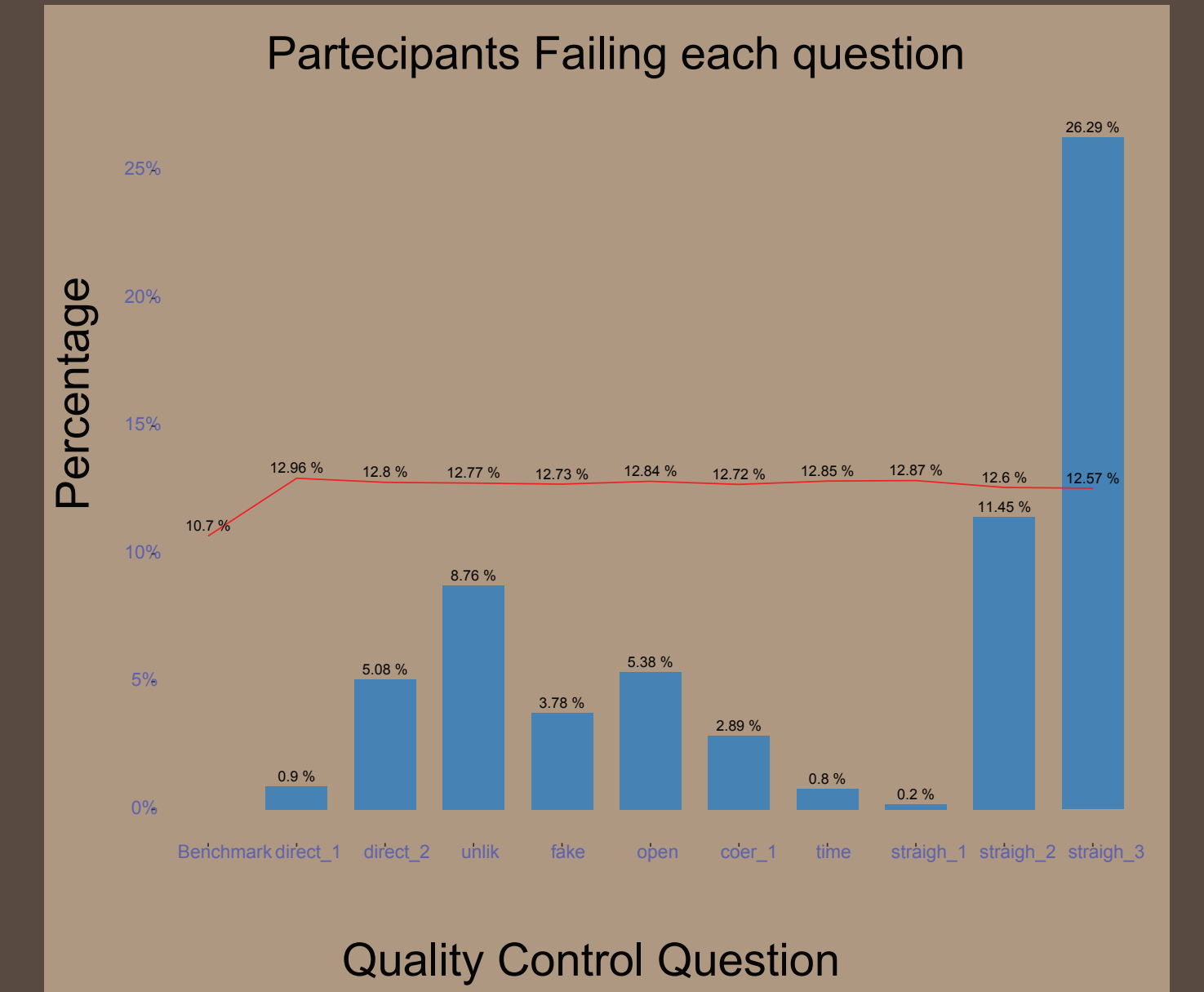
Suffering allergies

We analyzed the percentage of people suffering allergies in different sub-groups based on the number of failed checks. In the sub-group of those who fail 3 checks or 4+ checks, there is a considerable increase in those who claim to be allergic, maybe because the "Yes" option was the first one and they didn't read the question properly. This could be an indicator that those who fail 3 or more checks should be eliminated from the survey.

Suffering allergies

In this case we found the same results of the first data set for the group of those who fail 4+ checks while, for the group of those who fail 3 checks there is a considerable decrease in the number of people who claim to be allergic and the percentage is quite similar to the benchmark. Therefore it may be that it is not necessary to eliminate those who fail 3 controls but just those failing 4+.
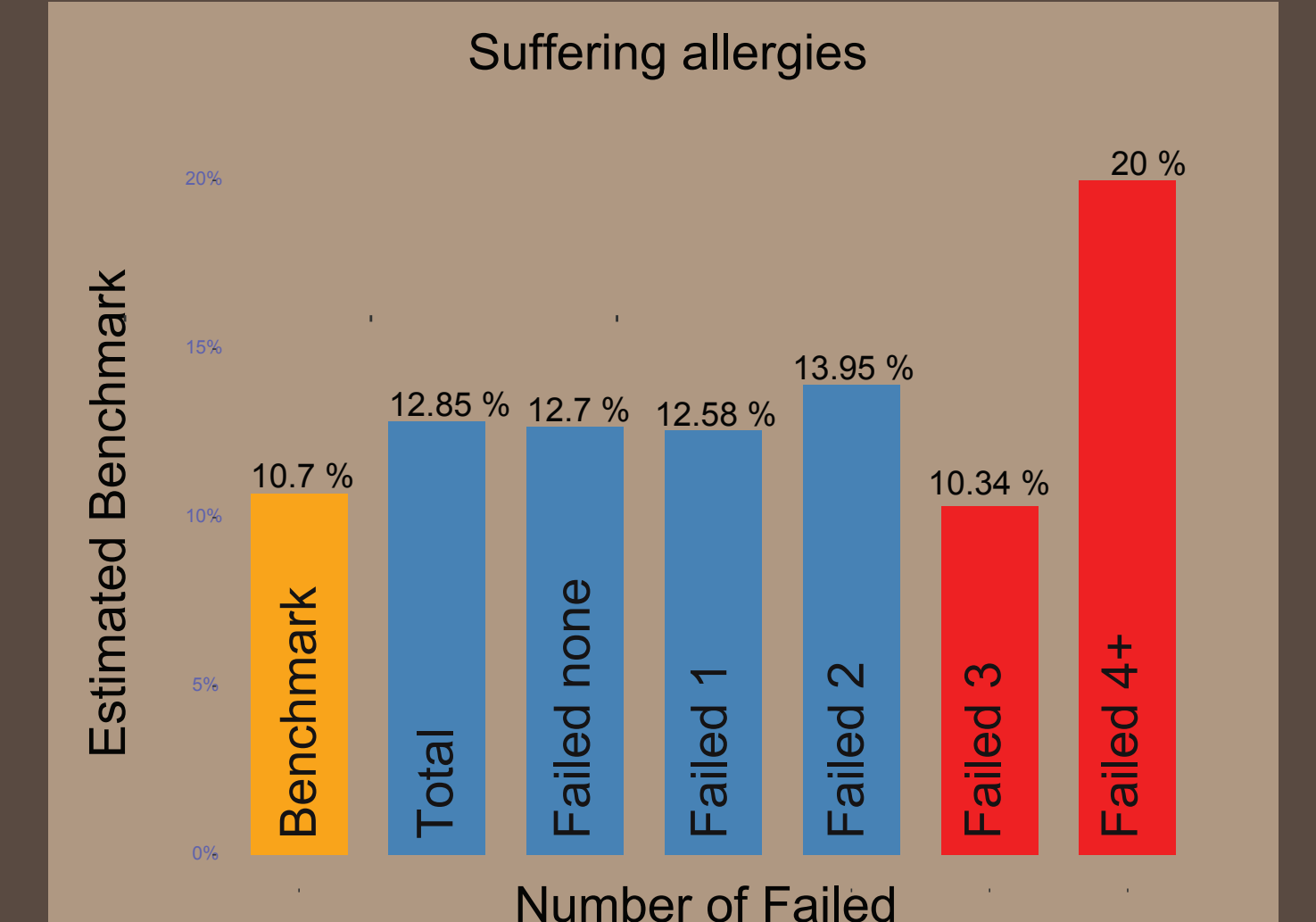
Suffering allergies

Finally, we compared the estimates of those suffering from allergies for different sub-groups based on the cumulate of the checks failed. There is no clear improvement of the estimate of the target variable under any circumstances. Eliminating those who fail 0 checks from the survey results in keeping only 57,6% of the sample, but this worsens the estimate so it is possible that we are eliminating too much information. On the other hand, those who fail 3 or more checks are too few to have an impact on the estimate of a dichotomous variable.
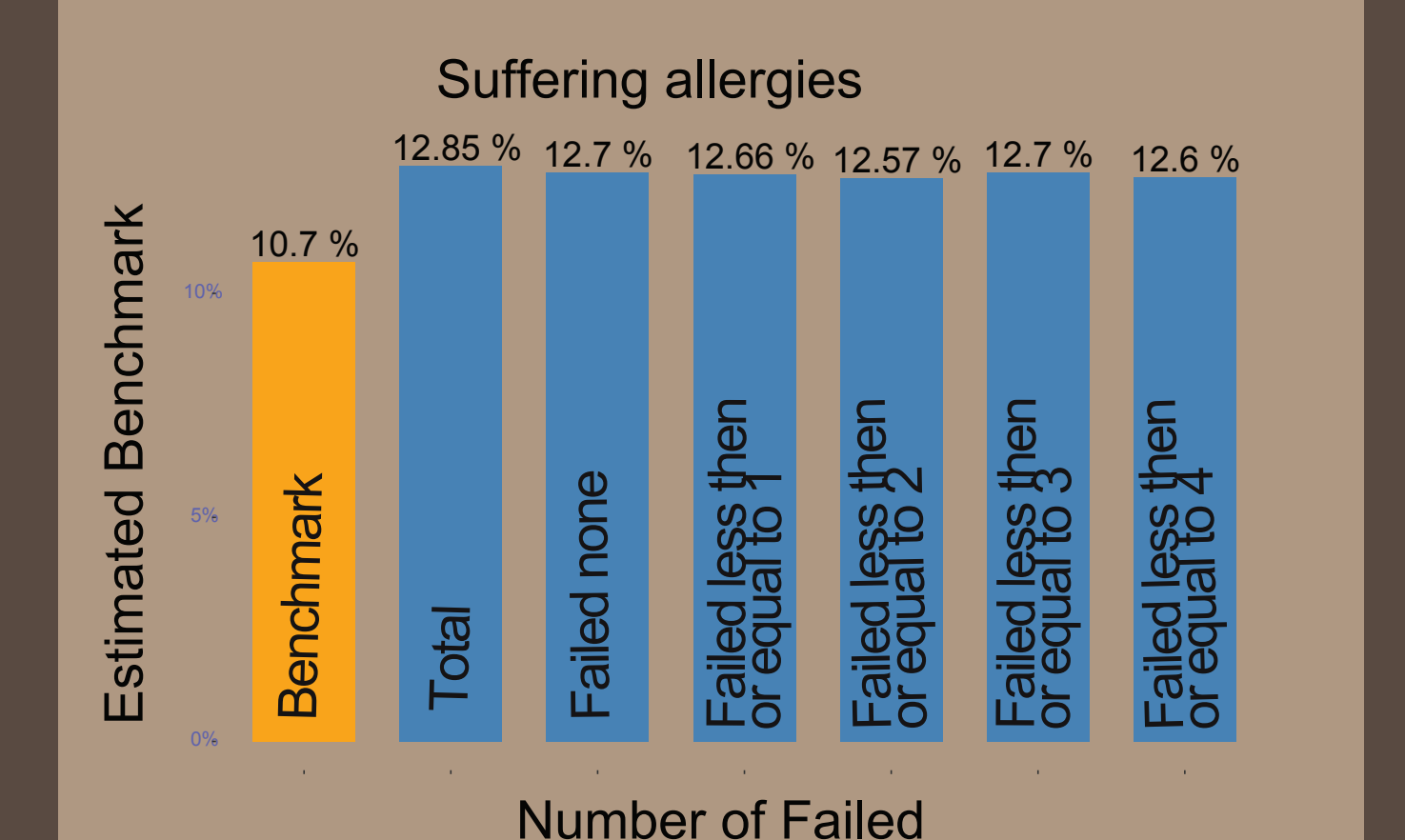
Suffering allergies

The results are quite comparable to that of the other data set.

Number of check failed

	Frequency	Percentage
0	618	57.6
1	329	30.7
2	74	6.9
3	28	2.6
4	12	1.1
5	7	.7
6	5	.5
Total	1073	100.0

Number of check failed

	Frequency	Percentage
0	567	56.5
1	302	30.1
2	86	8.6
3	29	2.9
4	8	.8
5	7	.7
6	5	.5
Total	1004	100.0

CONCLUSION:

The two biggest factors in determining the number of participants eliminated from a study are:

1) The way the quality control question is designed. We have found that the way in which direct instruction in question body and trap question with fake brands are designed, significantly changes the percentage of people who fail the checks. Also the position of the questions influences the failure of checks.
2) The number of quality control questions asked: Removing respondents who fail a single quality control question does not improve data quality. In our analysis, participants flagged for removal should fail at least 3 quality control measurements.