Biases in questionnaire construction: how much do they influence the answers given?

Afonso Quelhas, André Santos, Bárbara Araújo, Carolina Silva, Catarina Marques, Cláudia Oliveira, José Carlos Marafona, Laura Gonçalves, Luís Sousa, Marta Sanhá, Rita Costa, Tânia Costa

(fmup.21.2010@gmail.com)

Prof. Dr. Altamiro Costa Pereira, Prof. Dr.ª Cristina Santos e Luís Pinto, Turma 21

ABSTRACT

**Background:** Questionnaires are widely used among investigators, in order to collect information. However, questionnaires are susceptible to errors, both systematic and random. Since systematic errors, also called biases, can be minimized, many authors have described and catalogued some of them.

**Objectives:** To assess whether these so called ‘biases’, documented in previous research articles, really influence the answers given in questionnaires.

**Target population:** First year students of “Faculdade de Medicina da Universidade do Porto”.

**Methods:** Several biases were chosen from researched literature. Afterwards, supposedly biased questions and the respective non-biased ones were written and randomly assigned to two versions of the same questionnaire. That is, the biased question included on version A has its respective non-biased on version B. Both versions of the questionnaire were delivered to the students of nineteen different classes of the target population. Gathered data was analyzed using the statistical analysis software SPSS 19.0.

**Results:** There is statistical evidence that the described biases “Ambiguous question”, “Faking good”, “Belief vs. behaviour”, “Neutral opinion”, “Faulty scale” and “Technical jargon” can indeed lead to different answers. For the rest, no statistical evidence was found.

**Conclusion:** It was observed that some of those described biases did not change the respondents’ answer, while there is statistical evidence that others certainly altered them – these are actual biases.

**Key Words:** questionnaire, documented biases, influence on answers.
INTRODUCTION

“A good questionnaire is that one that works.”[1]

The purpose of any questionnaire is to collect information about a certain population. In fact, survey questionnaires have been one of the main primary data collection instruments in health research [2]. They also have proven to be of extreme importance in understanding human nature [3,4].

Building a questionnaire is not always easy and can be impracticable. Random and systematic errors, perpetrated during its execution, may interfere with the correct analysis of results.

Random errors are related to the variability inherent to a phenomenon that is being measured [5]. They directly affect questionnaires precision (its capacity to show the same results when submitted, under the same conditions, to repeated measurements). On the other hand, systematic errors might lead to an estimated statistic that does not correspond to the true value of the population [6]. This kind of errors affects questionnaires’ validity and they are commonly known as “biases”, which will be the point of focus of this work.

The standardized format of questionnaires makes them susceptible to biases. Consequently, the risk of manipulation is eminent. Therefore, in order to collect the most accurate data from respondents, investigators must understand and be able to prevent or at least minimize bias [7,8]. Here lies the importance of having a cautious preparation on the language used (taking into account the characteristics of the target population), on the accuracy of the questions made and even on the design of the questionnaire [9].

Biased questionnaires are common and many statistics and studies are analyzed based on false, or better, inaccurate information. Thus, many authors have described and catalogued biases frequently used in questionnaires [5,8,9,10]. The purpose of these catalogues is to inform investigators and, by this, help them to minimize the number of biases when building a questionnaire.

According to the article that serves as the primary basis of this research work [7]: there are three main categories of biases:

- those concerned with the way a question is built;
- the ones related to the way a questionnaire, itself, is designed;
- finally, those related to the way a questionnaire is administrated.

In the following paragraphs, it is presented a list of the most common biases along with a brief explanation of each one.
1. BIASES RELATED TO “QUESTION DESIGN”

- Problems with wording
  - *Ambiguous question*
    
    Leads respondents to understand wrongly the question and, therefore, answer a different question than the one that was intended.
  
  - *Technical jargon*
    
    Deals with the fact that technical terms may not be understood by all the respondents.
  
  - *Uncommon words*
    
    Uncommon and difficult words should be avoided in questionnaires in order to minimize misunderstandings. This kind of words should be replaced by simpler alternatives.

- Faulty scale
  - *Missing interval*
    
    Response choices are presented as a scale in which there are flaws or gaps omitted, so the scale does not fit all respondents and it can cause confusion.

- Leading questions
  - *Leading question*
    
    It has to do with the use of words that direct respondents towards a certain answer.

- Intrusiveness
  - *Sensitive questions*
    
    Respondents are more likely to refuse to answer this type of questions which focus on, for example, age, sexual orientation or other sensitive issue.

- Missing or inadequate data for intended purpose
  - *Belief vs behaviour*
    
    Deals with the fact that many respondents have certain beliefs that do not correspond to their own behavior. So, questions that ask the respondent about a belief (hypothetical) can yield quite different answers than questions that ask the respondent about his or her behaviors (personalized).
2. BIAS IN QUESTIONNAIRE DESIGN

- **Formatting problem**
  - *Horizontal response format*

  Questions designed horizontally can be misleading because they confuse respondents on which square to cross. If respondents want to answer a certain option, they would be undecided on whether to cross the square on the right or on the left. Choices of a non-biased questionnaire would be designed vertically, so there would not be any kind of mistake on the answer given.

3. ADMINISTRATION OF QUESTIONNAIRE

- **Neutral opinion options**

  The "neutral opinion" option allows respondents to state that they have no opinion or have not thought about a particular issue. It might be an easy way out of answering a certain question. The physical placement of the "undecided" category (at the midpoint of the scale) can change response patterns.

- **Respondent's conscious reaction**
  - *Faking good*

  Respondents may systematically alter questionnaire responses in the direction they perceive to be desired by the investigator. This kind of bias deals with the fact that, when people are confronted with certain types of questions, consciously, they tend to answer according to sociable acceptance. A different answer would be socially disapproved.

**RESEARCH QUESTION AND AIMS**

As mentioned before, many articles describe biases and, accordingly to their nature, they are inserted into a certain category. However, there are few evidences that what is described in the literature is really the cause of the deviation in results. Thus, this leads to a question: *Do these so called biases really influence the answers given in questionnaires?*

In order to obtain an answer to our research question, mentioned above, aims were established.

Firstly, check if the way one changes a normal question influences the given answers in general. Secondly, and more particularly, be able to sort out at what extent each of the types of biases, mentioned in introduction, can manipulate the results obtained. To sum up, it is our goal to learn about the influence of catalogued biases in questionnaires.
METHODS

In order to achieve our goals, it was decided that two versions of the same questionnaire should be elaborated and distributed to first year students of “Faculdade de Medicina da Universidade do Porto”.

Before starting to build those questionnaires, a detailed research was made to understand what biases would be more adequate to include, using various articles as reference. During this research, examples of so called “biased questionnaires” were also found. After that, the selection of documented biases started. They were chosen according to the interest they could pose within the outline of this research work.

A theme for the questionnaires was then chosen: “How do people of FMUP live? - The lifestyle of medical students”. Afterwards, the supposedly biased questions and the respective non-biased questions were written. We were encountered with the difficulty of creating a non-biased question.

The supposedly biased and respective non-biased questions were randomly assigned to versions A and B of the questionnaire. That is, the biased question included on questionnaire version A had its respective non-biased question on questionnaire version B. In version A, the biased questions are: 1, 2, 4, 7 and 10. In version B, they are: 4, 6, 7, 9 and 10 (see attached questionnaires). Therefore, half of the ten questions included in each version of the questionnaire are as impartial as possible and the other half contains a certain type of bias, according to literature.

In the next phase, questionnaires were delivered to nineteen different classes. The distribution of questionnaires version A or B was made during “Introdução à Medicina II” classes. Therefore, questionnaires were distributed to two classes at the same time, in such a way that participants could not know that two versions of the same questionnaires existed. After gathering all the answers given, supposedly biased questions were compared with the respective non-biased ones.

The analysis of the answers given was only possible because we consider that the odds of obtaining a certain answer is similar in both groups of each class. Our outcomes were the rates in the answers given to supposedly biased and unbiased questions. These rates will be compared to check if there are any significant differences. To analyse our results we used the statistical analysis software SPSS 19.0. The hypothesis test used to assess the influence of those supposedly biases on the answers given was chi-square test.

Below, it is presented the types of biases and questions (see questionnaire version A and B attached) included in the questionnaires along with a brief explanation of the purpose of each one of them.
QUESTION DESIGN

Problems with wording - Ambiguous question
It was expected that people answer ‘No’ in the supposedly biased question (A1), since normally they do not lend their tooth brush, but only their tooth paste – this question is ambiguous because the respondent does not know whether to consider the tooth brush or the tooth paste. In the unbiased questions (B1 e B2), they might answer differently in each one of them.

Missing or inadequate data for intended purpose - Belief vs behavior
If we are trying to evaluate if people practice physical exercise 3 times a week, we should not ask if they believe it is important, but if they actually do it – the answer might be different, according to the literature. (Refer to questions A5 and B6.)

Faulty scale - Missing interval
In the allegedly biased question (B9), the scale is incomplete, which might lead respondents to answer a false number or not answer at all.

Leading question - Leading question
The biased question (B10) makes people regard only ‘cinema’ as the only activity, not considering other possibilities. The answer in the supposedly unbiased one (A9) can be different.

Problems with wording - Uncommon word
If the respondent does not know the meaning of the words ‘reckon’ or ‘germane’, they might not answer this ‘biased’ question. (Refer to questions A4 and B5.)

Problems with wording - Technical jargon
The Wernicke’s area is the area of the brain you use to understand and process oral and written language [11]. Given that most of the respondents are not aware of this fact, it might lead to a high rate of non-answers in the supposedly biased question. (Refer to questions A10 and B11.)
QUESTIONNAIRE DESIGN

**Horizontal response format**

This type of scale format might create confusion when it comes to choosing the right box to sign their answer. (Refer to questions A6 and B7.)

ADMINISTRATION OF QUESTIONNAIRE

**Respondent’s conscious reaction - Faking good**

It is more socially acceptable to have a shower everyday, which might cause the respondents to answer ‘Yes’, despite the fact that it may not be true. (Refer to A3 and B4.)

**“Neutral opinion” options**

According to the literature, this kind of biases may lead people to say that they have no opinion about this issue, in order to save time or not have to think about it. (Refer to A7 and B8.)
RESULTS

Table 1. Absolute values and respective percentages obtained in each option used to verify the relation between supposedly biased question and respective unbiased question. Note that percentages are presented in row. Test chi-square was used to calculate p values. (Cut off for statistical significance: 0.05)

<table>
<thead>
<tr>
<th></th>
<th>Version A n=99</th>
<th>Version B n=96</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous question (A1/B1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>22 (25)</td>
<td>65 (75)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No</td>
<td>77 (73)</td>
<td>29 (27)</td>
<td></td>
</tr>
<tr>
<td>Ambiguous question (A1/B2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>22 (100)</td>
<td>0 (0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No</td>
<td>77 (45)</td>
<td>93 (55)</td>
<td></td>
</tr>
<tr>
<td>Sensitive question (A2/B3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>99 (52)</td>
<td>92 (48)</td>
<td>0.301</td>
</tr>
<tr>
<td>No</td>
<td>0 (0)</td>
<td>1 (100)</td>
<td></td>
</tr>
<tr>
<td>Faking good (A3/B4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than 6 days (A)/Yes(B)</td>
<td>67 (45)</td>
<td>82 (55)</td>
<td>0.001</td>
</tr>
<tr>
<td>Other options* (A)/No(B)</td>
<td>32 (73)</td>
<td>12 (27)</td>
<td></td>
</tr>
<tr>
<td>Sensitive question (A2/B3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>85 (48)</td>
<td>91 (52)</td>
<td>0.182</td>
</tr>
<tr>
<td>No</td>
<td>7 (70)</td>
<td>3 (30)</td>
<td></td>
</tr>
<tr>
<td>Horizontal response format (A6/B7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>4 (57)</td>
<td>3 (43)</td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>17 (61)</td>
<td>11 (39)</td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>55 (50)</td>
<td>56 (50)</td>
<td>0.463</td>
</tr>
<tr>
<td>Fair</td>
<td>14 (41)</td>
<td>20 (59)</td>
<td></td>
</tr>
<tr>
<td>Somehow deficient</td>
<td>5 (71)</td>
<td>2 (29)</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>2 (50)</td>
<td>2 (50)</td>
<td></td>
</tr>
<tr>
<td>Very poor</td>
<td>2 (100)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Neutral opinion (A7/B8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agree</td>
<td>49 (45)</td>
<td>59 (55)</td>
<td>0.031</td>
</tr>
<tr>
<td>Disagree</td>
<td>12 (27)</td>
<td>33 (73)</td>
<td></td>
</tr>
<tr>
<td>Faulty scale (A8/B9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 days</td>
<td>11 (23)</td>
<td>37 (77)</td>
<td>0.003</td>
</tr>
<tr>
<td>5 days</td>
<td>18 (26)</td>
<td>51 (74)</td>
<td></td>
</tr>
<tr>
<td>7 days</td>
<td>11 (65)</td>
<td>6 (35)</td>
<td></td>
</tr>
<tr>
<td>Leading question (A9/B10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>68 (50)</td>
<td>69 (50)</td>
<td>0.538</td>
</tr>
<tr>
<td>No</td>
<td>30 (55)</td>
<td>25 (45)</td>
<td></td>
</tr>
<tr>
<td>Technical jargon (A10/B11)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>53 (36)</td>
<td>93 (64)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No/Missing</td>
<td>46 (34)</td>
<td>3 (6)</td>
<td></td>
</tr>
</tbody>
</table>

*Other options: 5-6 days per week; 2-4 days per week; 1 day per week; I never have a shower.
For some tested biases it was assessed the relation between the two questions (supposedly biased question and respective unbiased question) comparing the option of one with that same option of the other one. Those tested biases were: “Ambiguous question”, “Sensitive question”, “Uncommon word”, “Belief vs. behavior”, “Horizontal response format”, “Leading question” and “Faulty scale”.

However, for biases like “Faking good”, “Neutral opinion”, Faulty scale” and “Technical jargon” that did not occur. Variables were recoded in order to compare questions in a way that allowed to assess biases’ influence in the answers given.

In questions A3 and B4, related to the bias “Faking good”, it was compared the rates obtained in options “Yes” and “More than 6 days”, corresponding respectively to questions B4 and A3. The rates obtained in “No” option (question B4) was compared with those obtained in all the other options of question A3.

As for the bias described as “Neutral opinion”, tested in questions A7 and B8, “Neutral opinion” rates were not included in the chi-square test performed for this case. Once both questions are equal in the two versions of the questionnaire, except for the neutral opinion option, the statistically significant difference observed between the two questions is due to the presence or absence of “Neutral opinion” option.

Concerning the bias “Technical jargon”, “No” and “Missing” rates were grouped. The rates obtained in both questions were compared. “Yes” option rates were also compared in both questions.

**DISCUSSION**

There is statistical evidence that the following described biases influence indeed respondent’s answers in questionnaires: “Ambiguous question”, “Faking good”, “Belief vs. Behavior”, “Neutral opinion”, “Faulty scale” and “Technical jargon”.

On the other hand, there is no evidence that “Sensitive question”, “ Uncommon word”, “Horizontal response format” and “Leading question” are indeed biases.

Encountered with this results, it is legitimate to affirm that:

- Two or more questions should not be fused into one once that creates confusion, leading to different answers (Ambiguous question).
- Words considered susceptible to hurt respondents sensitivity do not appear to influence respondents answers in questionnaires (Sensitive question).
- When investigators intend to ask about someone’s behavior it should be taken into attention that asking for someone’s behavior is different from asking his/her belief (Belief vs. behavior).
- What is considered more socially acceptable seems to have influence on participants answers (Faking good).
- The absence of a neutral opinion option appears to force people to choose, influencing the results obtained (Neutral opinion).
- The presence of a faulty scale seems to change the way participants respond to a question (Faulty scale).
- Using technical jargons in questions appears to influence respondents answers, since they do not understand what is being asked (Technical jargon).
- The application of a horizontal or vertical response format to a questionnaire seems to be indifferent (Horizontal response format).
- Giving examples of what it is intended to ask does not seem to set participants mind towards a certain answer (Leading question).
- Using uncommon words, such as “germane”, do not seem to influence respondents answers (Uncommon word).

However, there are some points that are worthy of discussion. Since online dictionaries are becoming more common, if participants do not understand technical jargon or any uncommon word it is possible that they check their meaning easily online. So the effect of the use of technical jargon might be dissolved in those cases.

As for the uncommon word used in this questionnaire, it might not have been the most adequate to use. Moreover, students had access to the internet while they were responding to the questionnaire, which might have influenced this result.

The present study is limited to a test to 9 biases described in literature. There are many others described whose influence has not been assessed through this study.

It is also important to notice that if applied this same questionnaire to a different population, for example, from to a different age group, results might not have been the same.

This is the first study described in literature that uses hypothesis tests to assess at what extent those so called biases really influence the answers given in questionnaires. According to C.K. Choi et al, in A Catalog of Biases in Questionnaires, all the supposedly biases, that were tested in this work, are assumed as factors of influence in responses to questionnaires. However, it has been observed that some of those described biases did not lead respondents to a certain answer (the description might be linked to prejudice), while there is statistical evidence that others are indeed biases.
REFERENCES

No âmbito da unidade curricular Introdução à Medicina, realizámos este questionário com o objectivo de avaliar o estilo de vida dos estudantes de Medicina do primeiro ano da FMUP. O presente questionário encontra-se em Inglês para facilitar o processamento de dados e inclusão destes no artigo final. Este questionário garante o seu anonimato.
Para cada uma das questões apresentadas, assinale com um (X) a opção de resposta que mais se adequa à sua opinião.

1. Do you lend your tooth paste or tooth brush to other people?
   Yes [ ]
   No [ ]

2. When you take a bath, do you wash your genital area?
   Yes [ ]
   No [ ]

3. How often do you have a shower in the week?
   More than 6 days per week [ ]
   5-6 days per week [ ]
   2-4 days per week [ ]
   1 day per week [ ]
   I never have a shower [ ]

4. Do you reckon that physical exercise is a germane aspect of your health?
   Yes [ ]
   No [ ]

5. Do you practice physical exercise at least 3 times a week?
   Yes [ ]
   No [ ]

6. Your eating habits are:
   Excellent [ ]
   Very good [ ]
   Good [ ]
   Fair [ ]
   Somehow deficient [ ]
   Poor [ ]
   Very poor [ ]
7. Do you agree that McDonalds is responsible for the increase of hypertension rates?
   Agree [ ]
   Neutral opinion [ ]
   Disagree [ ]

8. How many days per week do you dedicate to study anatomy?
   0 days [ ]
   1 day [ ]
   2 days [ ]
   3 days [ ]
   4 days [ ]
   5 days [ ]
   6 days [ ]
   7 days [ ]

9. Do you arrange activities with your friends every week?
   Yes [ ]
   No [ ]

10. Did you exercise your Wernicke’s area while responding to this questionnaire?
    Yes [ ]
    No [ ]

Obrigado pelo tempo dispensado,
       Turma 21
No âmbito da unidade curricular Introdução à Medicina, realizámos este questionário com o objectivo de avaliar o estilo de vida dos estudantes de Medicina do primeiro ano da FMUP. O presente questionário encontra-se em Inglês para facilitar o processamento de dados e inclusão destes no artigo final. Este questionário garante o seu anonimato.
Para cada uma das questões apresentadas, assinale com um (X) a opção de resposta que mais se adequa à sua opinião.

1. Do you lend your tooth paste to other people?
   Yes [ ]
   No [ ]

2. Do you lend your tooth brush to other people?
   Yes [ ]
   No [ ]

3. When you take a bath, do you wash your hair?
   Yes [ ]
   No [ ]

4. You have a shower every day, don´t you?
   Yes [ ]
   No [ ]

5. Do you consider that physical exercise is a relevant aspect of your health?
   Yes [ ]
   No [ ]

6. Do you think it is important to practice physical exercise at least 3 times a week?
   Yes [ ]
   No [ ]

7. Your eating habits are:
   Excellent [ ] Very Good [ ] Good [ ] Fair [ ] Somehow deficient [ ] Poor [ ] Very Poor [ ]

8. Do you agree that McDonalds is responsible for the increase of hypertension rates?
Agree [ ]
Disagree [ ]

9. How many days per week do you dedicate to study anatomy?
2 days [ ]
5 days [ ]
7 days [ ]

10. Do you arrange activities with your friends, such as cinema, every week?
Yes [ ]
No [ ]

11. Did you understand what you read in this questionnaire?
Yes [ ]
No [ ]

Obrigado pelo tempo dispensado,
Turma 21