A comparison of cigarette smokers recruited through the Internet or by mail

Jean-François Etter and Thomas V Perneger

Objectives

To compare smokers recruited by mail or through the Internet.

Methods

A questionnaire was mailed to 19,352 inhabitants of Switzerland in 1998, in an effort to enrol them in a smoking cessation trial. The same questionnaire was also available on the Internet. Furthermore, we mailed a survey to a representative sample (n = 1000) of the population of Geneva, Switzerland, in 1996. In this study, we compare three groups: 1027 smokers recruited through the Internet, 2961 volunteer trial participants recruited by mail (response rate 16%), and 211 smokers in the representative sample also recruited by mail (response rate 75%).

Results

Smokers self-recruited through the Internet were younger, more educated, more motivated to quit smoking and smoked more cigarettes per day than smokers in the other samples. Compared to trial participants, Internet participants had more negative attitudes towards smoking, higher self-efficacy scores, and were more addicted to tobacco. The strength of associations between smoking-related variables was similar in Internet and trial participants.

Conclusion

As expected, the three groups of smokers differed on several characteristics. However, bias in distributions of variables did not imply bias in associations between variables. Thus, Internet recruitment is a potentially useful method for analytical studies that focus on associations between variables.

Keywords

Smoking prevention and control, internet, mail surveys, bias

Accepted

17 October 2000

The Internet is a fast and cost-effective tool for data collection in epidemiological and medical research.1–6 Internet questionnaires are easy to design and to answer, and computer programs can provide real-time evaluation of answers and ensure that the data are complete and accurate before they are accepted.2 Internet surveys are also cheap and data entry errors by research assistants are eliminated.

Despite this potential, few published studies used health-related data collected on the Internet,3–6 or compared data collected on the Internet to similar data collected by mail.4 Concerns about selection bias, data quality, eligibility of participants, representativeness of samples or about the possibility of the same person being registered several times could explain the limited use of the Internet in health research. However, these hypotheses are largely untested.

We compared smokers who volunteered for a smoking cessation trial and were recruited by mail to smokers self-recruited through the Internet to receive smoking cessation counselling. Since both Internet participants and trial participants may differ from unselected smokers in the general population, we compared these two samples to a representative sample of smokers also surveyed by mail. We examined bias in distributions of smoking-related variables and in associations between variables.

Methods

Setting

Internet participants

Internet participants visited, between June 1998 and February 1999, a French-language Internet site on smoking cessation (www.stop-tabac.ch). They found the site either on search engines or from links on health-related web sites. Participants were invited to answer a 61-item questionnaire in order to obtain, a few seconds later, a smoking-cessation counselling letter tailored on their answers. Participants could choose whether or not they wanted their data to be archived, and they were informed that archived data would be used for statistical analyses.

Institute of Social and Preventive Medicine, University of Geneva, Switzerland.

Correspondence: Jean-François Etter, Institute of Social and Preventive Medicine, University of Geneva. CMU, Case Postale, CH-1211 Geneva 4, Switzerland. E-mail: etter@cmu.unige.ch Internet: www.stop-tabac.ch

521
**Trial volunteers recruited by mail**

A random list of 20,000 addresses was drawn from the official file of residents of the French-speaking part of Switzerland. These people (aged 18–60) received by mail a questionnaire and an invitation to participate in a randomized trial aimed at evaluating a smoking cessation counselling programme. This programme included a series of individually tailored counselling letters and stage-matched booklets (available on www.stop-tabac.ch). The questionnaire used to produce the counselling letters was the same in the Internet and trial samples. Only daily smokers were eligible for the trial. Those who were not eligible and smokers who declined participation were asked to transmit the questionnaire to any smoker they knew. Participants who received the questionnaire directly from us and those who received it indirectly, from an addressee, were similar on most demographic and smoking-related variables. Therefore, these respondents were grouped together. This questionnaire was mailed only once and no reminder mailings were sent out. Participants in the trial were informed that they would have to answer a follow-up questionnaire 6 months later.

**Representative population sample**

A questionnaire on smoking and alcohol prevention was mailed in 1996 to a representative (random) sample of 1000 Geneva residents aged 18–70 years, identified through the official resident registry. This questionnaire did not contain psychometric scales on smoking-related attitudes, self-efficacy or self-change strategies. Non-respondents received up to five reminder mailings and transmission of the questionnaire was not allowed.

**Group comparisons**

Because we knew of no published comparison between smokers recruited by mail or through the Internet, we did not specify a priori hypotheses on between-group differences. Rather, we compared groups according to variables that are relevant to smoking prevention.

We compared Internet-, trial- and representative samples on age, sex, number of school years, having children, number of cigarettes smoked per day, having made a 24-hour attempt to quit smoking in the past year, and stages of change. We classified current smokers in three stages of change: precontemplation (no intention to quit in the next 6 months), contemplation (seriously considering quitting in the next 6 months) or preparation (decided to quit in the next 30 days and has made an attempt to quit in the past 12 months).

The following variables were available for Internet and trial participants only:

**Data quality**

Multiple registration among Internet participants (in case of duplicate record, we accepted only the first record), proportion of obviously unreliable questionnaires (i.e. questionnaires that contained a majority of missing answers or a majority of contradictory answers) and, among valid questionnaires, proportion of missing answers.

**Components of the Transtheoretical Model of Behaviour Change**

Smoking-related self-efficacy, evaluation of the adverse effects of smoking, and the frequency of use of self-change strategies. Self-efficacy was assessed with a 12-item scale measuring two dimensions: confidence in one’s ability to refrain from smoking when facing internal stimuli (e.g. feeling depressed), and external stimuli (e.g. having a drink with friends). Evaluation of the negative effects of smoking was assessed with a 10-item scale.

**Involvement in behaviour change** was assessed with a 19-item scale measuring the frequency of use of five self-change strategies labelled ‘Risk assessment’, ‘Commitment to quit smoking’, ‘Taking control over the smoking habit’, ‘Coping with the temptation to smoke’ and ‘Helping relationships’. These multi-item scales were previously submitted to comprehensive validation tests. For a better interpretation of these psychometric scores, readers are referred to published data on the association between these scores and stages of change, level of dependence, and smoking status.

Level of addiction to tobacco was measured by the number of minutes between waking up and smoking the first cigarette of the day.

**Associations between variables**

Bias in the distributions of variables does not necessarily imply bias in associations between variables. We compared the strength of associations between smoking-related variables in Internet and trial participants. The Transtheoretical Model of Change was used as a framework for these comparisons. First, we compared the size of differences between smokers in the precontemplation stage and smokers in the contemplation or preparation stages, on scores of attitudes towards smoking, self-efficacy and self-change strategies. Second, we compared differences on these scores between light smokers (<15 cig./day) and heavy smokers (>20 cig./day).

**Statistical procedures**

Psychometric scales were expressed in standardized scores (mean = 50, SD = 10). We used $\chi^2$ tests to compare categorical variables, t-tests or ANOVA to compare continuous variables and to compute confidence intervals, and multivariate logistic regression to identify variables independently associated with Internet versus mail recruitment.

**Results**

**Participation**

**Internet**

About 15,000 people visited the Internet site between June 1998 and February 1999. The questionnaire was answered 1975 times, and 394 people (20%) asked us not to store their data, thus 1581 records were stored. We deleted 15 records that we estimated to be obviously unreliable (0.9%), 235 records of people registered twice (15%), and 62 follow-up assessments (4%). Of 1269 first assessment participants, 1027 daily smokers (81%) were included in subsequent analyses; occasional smokers were excluded.

**Trial survey**

Of 19,352 mailings sent to valid addresses, 3124 questionnaires were returned (16%), including 2961 by daily smokers (96%).

**Representative survey**

We received 751 questionnaires (75% of 1000 questionnaires sent out), from 211 smokers (28% of 751) and 540 non-smokers. Only smokers were included in this study.
Differences in distributions of variables

All participants in the representative sample and all but six participants (0.4%) in the trial lived in Switzerland. Internet participants lived in France (n = 337, 33%), Switzerland (n = 296, 29%), Canada (n = 149, 15%), Belgium (n = 31, 3%), and other countries (n = 214, 21%). Differences between Internet participants who lived in Switzerland or in other countries were small (data not shown). Compared to participants in the trial and to the representative sample, Internet participants were younger, more educated, less likely to have children; they smoked more cigarettes per day; more had attempted to quit smoking in the past year, and more were in the contemplation or preparation stages of change (Table 1).

Compared to trial participants, Internet participants were more frequently men; they were more addicted to tobacco (they smoked their first cigarette of the day earlier), and had higher scores on the ‘Adverse effects of smoking’ and ‘External stimuli’ self-efficacy scales. They were also more actively involved in changing their smoking behaviour, as shown by their higher scores on four of the five self-change strategies (Table 1).

In multivariate analysis, the following variables were independently associated with Internet recruitment: younger age, male sex, more education, contemplation or preparation stages (versus precontemplation), a more severe addiction to tobacco (i.e. smoking more cigarettes and smoking one’s first cigarette earlier in the morning), more frequent use of the

Table 1 Characteristics of daily smokers recruited through the Internet, daily smokers recruited by mail for inclusion in a smoking cessation trial and of a representative population sample of smokers in Geneva, Switzerland, 1996–1998

<table>
<thead>
<tr>
<th></th>
<th>Internet</th>
<th>Trial</th>
<th>Population sample</th>
<th>Difference Internet-trial</th>
<th>Difference Internet-population</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of smokers</td>
<td>1027</td>
<td>2961</td>
<td>211</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Average age (range) (years)</td>
<td>34.1</td>
<td>36.3</td>
<td>38.6</td>
<td>–2.2***</td>
<td>–4.5***</td>
</tr>
<tr>
<td>(15–74)</td>
<td>(15–79)</td>
<td>(18–71)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men (%)</td>
<td>56.5</td>
<td>47.7</td>
<td>57.3</td>
<td>8.8***</td>
<td>–0.8</td>
</tr>
<tr>
<td>Have children (%)</td>
<td>33.4</td>
<td>37.9</td>
<td>50.2</td>
<td>–4.5*</td>
<td>–16.8***</td>
</tr>
<tr>
<td>School years</td>
<td>15.5</td>
<td>14.2</td>
<td>12.7</td>
<td>1.3***</td>
<td>2.8***</td>
</tr>
<tr>
<td>Cigarettes per day</td>
<td>22.6</td>
<td>19.8</td>
<td>16.5</td>
<td>2.7***</td>
<td>6.1***</td>
</tr>
<tr>
<td>Minutes to first cigarette</td>
<td>46.9</td>
<td>70.0</td>
<td>–</td>
<td>–23.1***</td>
<td>–</td>
</tr>
<tr>
<td>Quit attempt in past year (%)</td>
<td>45.0</td>
<td>40.7</td>
<td>29.8</td>
<td>4.3*</td>
<td>15.2***</td>
</tr>
</tbody>
</table>

Stage of change

<table>
<thead>
<tr>
<th></th>
<th>Internet</th>
<th>Trial</th>
<th>Difference</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precontemplation (%)</td>
<td>12.1</td>
<td>41.2</td>
<td>–29.1</td>
<td>–61.5</td>
</tr>
<tr>
<td>Contemplation (%)</td>
<td>64.4</td>
<td>53.6</td>
<td>10.8</td>
<td>42.1</td>
</tr>
<tr>
<td>Preparation (%)</td>
<td>23.4</td>
<td>5.2</td>
<td>18.2</td>
<td>19.3</td>
</tr>
<tr>
<td>Adverse effects of smoking</td>
<td>52.6</td>
<td>49.1</td>
<td>3.5***</td>
<td>–</td>
</tr>
</tbody>
</table>

Self-efficacy

<table>
<thead>
<tr>
<th></th>
<th>Internet</th>
<th>Trial</th>
<th>Difference</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal</td>
<td>49.9</td>
<td>50.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>External</td>
<td>51.0</td>
<td>49.7</td>
<td>1.3***</td>
<td>–</td>
</tr>
</tbody>
</table>

Self-change strategies

<table>
<thead>
<tr>
<th></th>
<th>Internet</th>
<th>Trial</th>
<th>Difference</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk assessment</td>
<td>52.9</td>
<td>49.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Helping relationships</td>
<td>53.9</td>
<td>48.8</td>
<td>4.6**</td>
<td>–</td>
</tr>
<tr>
<td>Coping with temptation</td>
<td>51.1</td>
<td>49.6</td>
<td>1.5***</td>
<td>–</td>
</tr>
<tr>
<td>Commitment to quit</td>
<td>53.9</td>
<td>48.7</td>
<td>5.3***</td>
<td>–</td>
</tr>
<tr>
<td>Taking control</td>
<td>50.4</td>
<td>49.9</td>
<td>0.5</td>
<td>–</td>
</tr>
</tbody>
</table>

Missing answers (%)   | 0.8      | 0.8   | –           | 0          |

* P < 0.05; *** P < 0.001.

Differences in associations between variables

Between-stage differences in attitudes, self-efficacy and the use of self-change strategies were similar in Internet and trial participants (Table 3). Differences between light and heavy

Table 2 Odds ratios of belonging to the Internet sample versus trial sample, from a multivariate logistic regression model. Geneva, Switzerland, 1998

| Age (per 10 years)          | 0.78 | 0.72, 0.83 |
| School years (per year)     | 1.13 | 1.10, 1.16 |
| Cigarettes per day (per 10 cigarettes) | 1.22 | 1.13, 1.32 |
| Time to first cigarette (per 10 min.) | 0.97 | 0.96, 0.98 |
| Contemplation or preparation stage of change (versus precontemplation) | 3.48 | 2.71, 4.46 |
| Sex (men versus women)      | 1.15 | 1.01, 1.31 |
| Self-efficacy (SD units)    | 1.15 | 1.06, 1.25 |
| Use of self-change strategies |     |       |
| Helping relationships (SD units) | 1.19 | 1.09, 1.30 |
| Commitment to quit smoking (SD units) | 1.20 | 1.07, 1.34 |
smokers were also similar in both groups, except for small differences in the ‘Internal stimuli’ self-efficacy score and in the ‘Helping relationships’ score (Table 3). Thus, only 2 of 16 between-group comparisons in the strength of associations between variables, compared to 14 of 17 comparisons in the distributions of variables, differed significantly between the Internet and trial samples.

**Discussion**

A selection bias towards young educated males was observed in smokers self-recruited through the Internet for receiving smoking cessation advice, compared to a representative sample of smokers, and compared to smokers recruited by mail for inclusion in a smoking cessation trial. Participation in the Internet survey required computer literacy and access to a computer connected to the Internet. As use of the Internet is not yet widespread, the Internet sample may be biased towards relatively privileged ‘innovators’. This type of bias may decrease over time.

Because they were recruited for a smoking cessation intervention, both Internet and trial participants were more motivated to quit smoking than smokers in the representative sample. Internet participants were more motivated to quit and were more addicted to tobacco than trial participants. There are several possible reasons for this difference. The first is information bias: the medium (screen or paper) may influence the answers. Then, there are several types of selection bias. One stems from having access to Internet technology, as stated above. Another is due to the active versus passive mode of recruitment: Internet participants actively sought our web site, whereas trial participants received the questionnaire without having requested it. A further possible source of selection bias is that trial participants had to accept the requirements of a randomized clinical trial, including the risk of being allocated to the control arm, and the necessity of providing follow-up information 6 months later. These possible reasons for differences between Internet and trial participants are confounded in our study and their respective contributions cannot be distinguished.

The trial sample also differed from the representative sample on several variables, in part because the number of mailings, and hence the participation rate, were not similar in the two surveys. More importantly, these differences confirm that smokers who volunteer for smoking cessation studies differ from smokers in the general population.

In contrast with the numerous differences in descriptive statistics, the strength of associations between smoking-related variables was similar in the Internet and trial samples. This is an important result, showing that bias in distributions of variables does not imply bias in associations between variables. These results suggest that Internet recruitment is potentially useful for analytical studies that are focused on associations between variables. However, this finding needs to be replicated, as it may not apply to variables other than those we measured, or to other populations.

The quality of data collected through the Internet was comparable to the quality of data collected by mail. Few unreliable records (<1%) were found in the Internet database and participants who were registered twice were easily identified. All data collected through the Internet are time-stamped and identified with the code of the participant’s computer. This information can be used to delete duplicate responses.

The possibility of identifying the respondents raises the issue of confidentiality of data collected through the Internet. Researchers using the Internet should make sure that study procedures respect regulations on data protection, and that survey participants provide informed consent. In particular, participants should be informed that their answers are stored, they should be told what will be done with their data, and they should be given the possibility of refusing storage. Failure to respect principles of informed consent and data protection may generate mistrust, which in turn may negatively affect future epidemiological research on the Internet.

Collecting data on the Internet is feasible and can provide data of good quality and large samples (our database currently...
includes 30,000 people, still growing). Such large databases allow for innovative analyses of narrowly defined subgroups of participants. If participants are given an identification code, follow-up studies can be conducted on the Internet.

Smokers recruited through the Internet differ from smokers in the general population, and so Internet recruitment cannot be used to describe the characteristics of populations other than Internet users. Nevertheless, our results suggest that the Internet may be a cost-effective method of data collection for analytical studies that assess associations between variables.

Acknowledgements

This research was supported by grants from the Swiss National Science Foundation to Dr Etter (32–47122–96, 3233–054994.98 and 3200–055141.98), by the Swiss Cancer League, the Swiss Federal Office of Public Health, Health Authorities of the Cantons of Geneva and Jura, the Geneva Cancer League and the Swiss Foundation for Health Promotion.

KEY MESSAGES

- Collecting epidemiological data on the Internet is feasible and cost-effective, and can provide data of good quality and large samples.
- Smokers recruited through the Internet differed from smokers recruited by mail on several characteristics. However, bias in distributions of variables did not imply bias in associations between smoking-related variables.
- Internet recruitment is a potentially useful method for analytical studies that focus on associations between variables.

References