Title: Socioeconomic inequalities in smoking habits are still increasing in Italy

Authors:

Giuseppe Verlato (giuseppe.verlato@univr.it)
Simone Accordini (simone.accordini@univr.it)
Giang Nguyen (giang.nguyen@univr.it)
Pierpaolo Marchetti (pierpaolo.marchetti@univr.it)
Lucia Cazzoletti (czzlcu44@univr.it)
Marcello Ferrari (marcello.ferrari@univr.it)
Leonardo Antonicelli (l.antonicelli@ospedaliriunits.marce.it)
Francesco Attene (francesco.attena@unina2.it)
Valeria Bellisario (Valeria.bellisario@unito.it)
Roberto Bono (roberto.bono@unito.it)
Lamberto Briziarelli (lbrizi@unipg.it)
Lucio Casali (lcasali@unipg.it)
Angelo Guido Corsico (corsico@unipv.it)
Alessandro Fois (FoIs.Ale@libero.it)
MariaGrazia Panico (grazia.panico@tin.it)
Pavilio Piccioni (ppiccioni@qubisoft.it)
Pietro Pirina (pirina@uniss.it)
Simona Villani (svillani@unipv.it)
Gabriele Nicolini (G.Nicolini@chiesigroup.com)
Roberto de Marco (roberto.demarco@univr.it)

Version: 3
Date: 15 July 2014

Author's response to reviews: see over
Dear Editors,

I herewith re-submit the revised version of the manuscript “Socioeconomic inequalities in smoking habits are still increasing in Italy” as an Original Article for consideration for publication in BMC Public Health.

The manuscript was modified according to the reviewers’ suggestions. Point-by-point response to the concerns raised by the reviewers is given in the next pages.

Hoping that the revised manuscript could be suited for publication in BMC Public Health, we look forward to hearing from you.

Best wishes

Giuseppe Verlato

Sincerely yours,

[Signature]

Prof. Giuseppe Verlato

Sezione di Epidemiologia e Statistica Medica - Dipartimento di Medicina e Sanità Pubblica - Istituti Biologici 2 - Strada Le Grazie, 8 - 37134 Verona - Italy

Tel. +39 045 8027628 - Fax: +39 045 8027154 - E-mail: giuseppe.verlato@univr.it
Responses to reviewers

Manuscript_1956499488128610: Socioeconomic inequalities in smoking habits are still increasing in Italy

REVIEWER 1:

The paper is interesting and the topic covered is important. It is well written and easy to read. There are however some issues that need to be tackled or discussed more. Here I summarize them in two broad points.

1. The paper stresses the role of education as a main driver for socioeconomic inequality in smoking habits. This is a plausible result and one that is in line with most of the previous literature. At the same time, authors state that education is a better predictor than occupation. This statement is probably too strong. Education and occupation are typically very much related (in normal circumstances, social scientists agree that education generally determines occupation) and it is thus difficult to disentangle the effect of the former from the latter. A cross-sectional dataset in this sense is not suited for the purpose. One possible way could be to exploit an exogenous shock on employment, such as, for example, the recent economic crisis.

Reply: Indeed there is a strong collinearity between education and occupation. As shown in the next table, in GEIRD cross-sectional study people with primary school education were mainly workmen, housewives and unemployed; people who had attended also the secondary school were mainly workmen and clerks; people with high school diploma were mainly clerks and students, while graduated people were mainly clerks and freelancers. We acknowledged this strong relation in the Discussion section (lines 308-310).

Education is more stable in adults while occupational status can change as a consequence of economic crisis.

As reported in the Results section, a model including only education minimized the AIC with respect to a model considering only occupation. The difference was substantial when evaluating risk factors for smoking initiation (13673 vs. 13777) but only minor when addressing smoking cessation (5674 vs. 5678). Anyway according to your suggestion, we attenuated the importance of this finding, removing the comparison between education and occupation from the Conclusions and the main results.
The paper, however, does not cover the period of the economic crisis in Italy (2011 onward) and in any case never mention the issue. However, as shown by Gallus et al 2013 (PMID: 23956058), the rapid increase in unemployment does explain a lot in the changes of smoking trends in the US: occupation was a particularly important predicting variable of smoking prevalence in the aftermath of the crisis. Although authors cannot do much about these issues, they need to discuss more carefully: a) the link between education and occupation, b) the representativeness of the results after the economic crisis hit the country. These points should be added in the discussion.

Reply: Indeed the pattern of occupation has remarkably changed after the economic crisis: the proportion of unemployed was 6.5% in our survey, which is compatible with national Italian data in 2008 (6.7%), while it has peaked (we hope !) at 13.6% in the first trimester 2014 (http://dati.istat.it/Index.aspx?DataSetCode=DCCV_TAXDISOCCU). We added this information at the end of the Discussion and we tried to forecast the impact of economic crisis on smoking habits by taking into account the paper by Gallus et al (2013) (lines 331-338).

2. The statistical analysis is not clear.

**Reply:** Selection bias is a very important issue in epidemiological survey, as in mail-phone surveys current smokers tend to be late responders with respect to never-smokers. This kind of bias has been reported in smoking studies from all over the world [USA, Oakes 1973; Norway, Bakke 1990; Italy, de Marco 1994, Verlato 2010; Spain, Galobardes 1998; Sweden, Ronmark 1999; Finland, Kotaniemi 2001]. For instance current smokers were 34% in early responders versus 50.3% in late responders in a Finnish study [Kotaniemi 2001], 29.2% versus 38% in an Italian study [Verlato 2010], and 32.9% versus 43.1% in an American study [Oakes 1973]. In particular, heavy smokers primarily appear to delay their response [Seltzer 1975; Verlato 2010]. On the other hand, ex-smokers tend to be early responders in most [Oakes 1973; Verlato 2010] but not all studies [Galobardes 1998].

Hence, when response percentage is low, the prevalence of ex-smokers is overestimated, while the proportion of current smokers is underestimated [de Marco 1994; Verlato 2010], and the decreasing response to epidemiological surveys [Frank 2007; de Marco 2012; Bjerg 2012; Koshy 2012] can amplify this bias.

b. Page 7: Can authors explain better what they do to counteract the selection bias and why those procedures would solve the problem?

**Reply:** To correct for non-response bias during data analysis, several methods have been proposed [Drane 1991; Tennant 1991; de Marco 1994; Verlato 1994; de Marco 1995], which usually rely on the assumption that non-responders are similar to late responders. However, the effectiveness of these correction methods is the highest when response percentage is at least 60%, as the trend in prevalence across subsequent contacts is not constant, and hence not fully predictable [Verlato 2010]. Recently we adopted a new method, which takes into account both the type of contact (mail vs phone) and the promptness to respond (percentile rank of cumulative response) [de Marco 2012]. For instance, if in one center 3000 people are contacted and 1800 respond, the percentile rank of the last responder is 1800/3001 = 59.98%.

**References**


c. If authors consider the merging of two different dataset as a “repeated cross-section”, why don’t they treat the new dataset as such? In other words, in the multilevel analysis, why don’t authors provide a formal test of the differences in smoking behaviours in specific subgroups? For example, is the smoking prevalence in the 30-39 yo in 1998-2000 significantly lower than the smoking prevalence in the same group in 2007-2010? At the moment, the analysis only allows to conclude that prevalence changes according to the age class.

Reply: We disagree. In the first version of the manuscript, when commenting Table 3 (now Table 4) at the end of the Results section, we noticed that: “The declining trend varied as a function of gender (time-sex interaction: p=0.029), age class (time-age interaction: p<0.001) and occupation (time-occupation interaction: p=0.047): the time-related decrease in ever smoking was slightly larger in women than in men, and it was particularly pronounced in people aged >=40 years, while being not significant in people younger than 30 years. As regards occupation, the declining trend was particularly evident in housewives, managers, businessmen and freelancers, while starting smoking became even more common among unemployed.”

For instance as regards the 30-39 year age class, from Table 3 (now Table 4) it is apparent that the Odds Ratio of being an ever-smoker was 1.26 (1.14-1.40) in 1998-2000 and 1.00 (0.88-1.12) in 2007-2010. In both cases the reference category is the 20-29 year age class in 1998-2000. So we can appreciate that ever smoking decreased over time, but as outlined by the Reviewer a formal test of significance is lacking. Anyway we chose to perform a single test on the whole sample by evaluating significance of the interaction term, rather than testing significance of variation in each subgroup to avoid multiple testing bias.

Following the Reviewer’s suggestion, at the end of the Method section (lines 173-174) we clarified the purpose of testing the interactions as follows: “Significance of the interactions between time and either sex, age class, or occupation was also tested to verify whether temporal trends in smoking habits varied as a function of the latter variables”.

d. Do different groups (especially occupational) become more relevant than others in explaining smoking prevalence? In this sense, and taking in consideration point c. above, using a logit/probit model of the form:

Pr(S)=bX+…+a1*Occupation_y_all_years+a2*Occupation_y_2007-2010
(where S is smoking and X a series of covariates) would allow authors to understand whether the role of occupation_y is different in the period 2007-2010 (i.e. if a2 is significantly different from 0) and in which direction.

Reply: We used a logistic model where the link function is the logit = \log(\text{odds}) = \log[p/(1-p)] [McCullagh P and Nelder JA, Generalized linear model, 2nd edition, 1989, Chapman & Hall]. Hence in our model:

$$\text{Logit } (S) = bX + \ldots + a1*\text{Occupation}_y_{\text{all\_years}} + a2*\text{Occupation}_y_{2007-2010}$$

We had already investigated the point raised by the reviewer: 1) As shown in Table 3 (now Table 4) the interaction between occupation and time was significant (p=0.047), i.e. the effect of occupation on smoking initiation significantly changed over time; 2) In the Result section we had remarked that “the declining trend (in smoking habits) was particularly evident in housewives, managers, businessmen and free-lancers, while starting smoking became even more common among unemployed.”

Indeed we used a slightly different approach in this respect: 1) we reported the odds ratio = \exp(\text{coeff.}) rather than the coefficient itself, as the odds ratio is more easy to understand for the average reader than a model coefficient; 2) we gave the overall significance of the interaction term rather than the significance for each of the 17 odds ratios, to avoid multiple testing bias. The 95% confidence interval of each odds ratio allows to appreciate whether the odds ratio is significantly different from one, i.e. from the reference category (clerks in 1998-2000). Indeed we did not present significance of temporal changes in each occupational category, but this is not easy to perform without incurring in multiple testing bias.

e. How comparable are ISAYA data with GERID data? In order to consider the pool of both datasets as a repeated cross-section, authors need to discuss more how comparable they are. If they are not comparable in terms of smoking prevalence, maybe they are in terms of covariates, in which case authors could still understand what is the relative role of covariate X and how it changed in time (point d).

Reply: ISAYA and GEIRD were multicentre cross-sectional surveys on respiratory diseases, carried out by the same research team on random samples of young adults, using the same design. In both studies, random samples of 3,000 subjects aged 20–44 (men/women ratio = 1) were selected in each centre from the general population, using the local health authority registers. A screening questionnaire was mailed to each subject up to three times and then administered by telephone in case of non-response.
As stated in the Methods-Questionnaire section “The screening questionnaire, used in GEIRD, was the same questionnaire used in ISAYA [32] with the addition of questions on education level, outdoor exposure, history of asthma, rhinitis, chronic bronchitis and eczema, and life impairment.”

For these reasons we are confident to confirm that ISAYA and GEIRD surveys were comparable.

In the Methods section (lines 122-126) we added the following sentence “Of note, ISAYA and GEIRD were multicentre cross-sectional surveys on respiratory diseases, carried out by the same research team on random samples of young adults, using the same design, sampling strategy and questions on smoking habits. In the 4 centres participating in both studies the number of participants was 8931 in the first survey and 5162 in the second one.”

We presented some information on the distribution of covariates in GEIRD and ISAYA (Table 3 of the new version of the manuscript).

<table>
<thead>
<tr>
<th></th>
<th>ISAYA (n=8931)</th>
<th>GEIRD (n=5162)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centres</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Verona</td>
<td>2166 (24.3%)</td>
<td>1746 (33.8%)</td>
<td></td>
</tr>
<tr>
<td>Pavia</td>
<td>2444 (27.4%)</td>
<td>966 (18.7%)</td>
<td></td>
</tr>
<tr>
<td>Turin</td>
<td>2266 (25.4%)</td>
<td>1205 (23.3%)</td>
<td></td>
</tr>
<tr>
<td>Sassari</td>
<td>2055 (23.0%)</td>
<td>1245 (24.1%)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Men</td>
<td>4439 (49.7%)</td>
<td>2397 (46.4%)</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>4492 (50.3%)</td>
<td>2765 (53.6%)</td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Clerk</td>
<td>3444 (38.7%)</td>
<td>2114 (41.3%)</td>
<td></td>
</tr>
<tr>
<td>Housewife</td>
<td>786 (8.8%)</td>
<td>352 (6.9%)</td>
<td></td>
</tr>
<tr>
<td>Manager</td>
<td>151 (1.7%)</td>
<td>90 (1.8%)</td>
<td></td>
</tr>
<tr>
<td>Businessman</td>
<td>268 (3.0%)</td>
<td>153 (3.0%)</td>
<td></td>
</tr>
<tr>
<td>Free-lancer</td>
<td>901 (10.1%)</td>
<td>570 (11.1%)</td>
<td></td>
</tr>
<tr>
<td>Workman</td>
<td>1381 (15.5%)</td>
<td>777 (15.2%)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>564 (6.3%)</td>
<td>316 (6.2%)</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>864 (9.7%)</td>
<td>549 (10.7%)</td>
<td></td>
</tr>
</tbody>
</table>
f. It would be nice to see a graph showing the results presented in the first paragraph of the “Results” section.

**Reply:** *These results are already presented in Table 1.*
REVIEWER 2:

1. In the Methods section and in the Abstract, age range is reported to be 20-44, while in Tables and in the main text, age range is 20-47. Please, correct the mistake.

Reply: The study design of GEIRD, ISAYA and ECRHS (European Community Respiratory Health Survey) included 1st, 2nd, 3rd postal waves and one phone contact. Indeed non-responders to the 1st postal wave were contacted again, first by mail and then by phone, to enhance response percentage and reduce non-response bias. For this reason, the screening phase lasted about two years, from sample selection to the last phone contact. Hence, the sample was selected from the general population aged 20-44 years, but age at interview ranged from 20 to 47 years. In the Methods – Study Design section (lines 115-117) we added a sentence to explain this discrepancy.

2. Please, add more information about the ISAYA study. Authors included a relevant reference, but it could be helpful to include at least the number of study participants.

Reply: We added a table (Table 3) presenting the covariate distribution in ISAYA and GEIRD. In the 4 centres participating in both studies the number of participants was 8931 in the first survey and 5162 in the second one.

We also added the following sentences to the 2nd paragraph of the Study design (lines 122-126): “Of note, ISAYA and GEIRD were multicentre cross-sectional surveys on respiratory diseases, carried out by the same research team on random samples of young adults, using the same design, sampling strategy and questions on smoking habits. In the 4 centres participating in both studies the number of participants was 8931 in the first survey and 5162 in the second one.”

3. Please, report the overall prevalence of current, never and ex-smokers, since only stratified results are provided in the Results section and in Tables.

Reply: We added the following sentence to the 1st paragraph of the Results section (lines 179-180): “Of these, current smokers were 2854 (27.7%) and ex-smokers 1662 (16.2%).”

4. It is not clear to me the choice of including in the analyses the number pack years at age 30.

Reply: We wanted to give a measure of the overall impact of smoking intensity (cigarettes/day) and smoking duration on smoking exposure (pack-years). For instance, in the 2nd paragraph of the Results section, we noted that “smokers with lower education had started smoking at an earlier age and consumed more cigarettes per day than smokers with higher education. As a
consequence, cumulative smoking exposure at the age of 30 years was nearly doubled in current smokers with only primary school license with respect to those with a university degree.”

We chose to compute pack-years at 30 years, because we tried to make a compromise between age at starting smoking and actual age of current smokers. Indeed in GEIRD 99% of current smokers in had started smoking by 30 years, while 70% of smokers were aged 30 years and over.

5. Tables and Figures should be improved.

- For example, authors should add more information to the caption and/or title of tables and figures, so that the reader doesn't need to look at the Methods and to the Results to understand, for instance, which study results refer to (GEIRD or ISAYA, in Figure 1), the year of the study, the adjustment of the model (Figure 1).

Reply: According to the Reviewer’s suggestion, we added the phrase “in the GEIRD cross-sectional study, performed in 2007-2010” to the legend of Figure 1. In the legend of all Tables we mentioned the study and the study period.

- Authors reported that information on smoking is available in 10,289 subjects, thus the analyses are based on this number of participants. However, in Table 1, the sum of the number of participants in strata of different socio-demographic characteristics is different by 10,289 (e.g., 4903 men + 5383 women=10,286). Please, specify, for example in a footnote, whether there are some missing values in sex, age and other characteristics. This happens also in Table 2 with current smokers only.

Reply: According to the Reviewer’s suggestion, we added the following caption to Table 1: “information on sex, age, education, occupation, site of residence was missing respectively in 3, 28, 59, 48, 1922 subjects. Of note, the Ancona sample (n=1866) was not asked about their site of residence.”. Among current smokers (Table 2) the same information was missing in 1, 7, 15, 13, 501 subjects respectively. Rather than presenting again missing numbers, we simply reported in a footnote that the Ancona sample was not asked about their site of residence.

- Are p values in Table 1 for the comparison only between current and ex-smokers, or are they for the comparison between never, current and ex-smokers (it would change the number of total subjects on which the test is based and the number of the degrees of freedom)? If this latter case is the one authors used, maybe it could be useful to include the column with the prevalence also of never smokers, in order to let the reader better understand.
**Reply:** The p values in the final column of Table 1 are for the comparison among never, current and ex-smokers. As suggested by the reviewer, we have modified the 2nd column of Table 1, even if this increases the table complexity.

- Table 2 provides incomplete results (e.g. interquartile range for cigarettes/day in men).

**Reply:** We have corrected the table according to the reviewer’ suggestion.

- Table 3: are p-values for age and occupation for the trend? If this is the case, please, specify it in a footnote.

**Reply:** P-values for age and occupation are not for trend. Indeed we preferred not to make a priori linear assumption as regards the association between age and either smoking initiation or cessation. Moreover, we considered occupation as nominal variable rather than precisely ranking different professions according to socio-economic status: for instance, which is the highest socioeconomic status, freelancers, businessmen or managers?

Quality of written English: Needs some language corrections before being published

**Reply:** We revised the English style.