Author's response to reviews

Title: Patients in long-term maintenance therapy for drug use in Italy: analysis of some parameters of social integration and serological status for infectious diseases in a cohort of 1091 patients.

Authors:

Gianluca Quaglio (gianluca.quaglio@azosp.vr.it)
Fabio Lugoboni (fabio.lugoboni@azosp.vr.it)
Cristian Pattaro (cristian@biometria.univr.it)
Gics Investigators (medicina.dipendenze@azosp.vr.it)
Linda Montanari (Linda.Montanari@emcdda.org)
Alessandro Lechi (alessandro.lechi@univr.it)
Paolo Mezzelani (paolo.mezzelani@univr.it)
Don CD Jarlais (dcdesjarla@aol.com)

Version: 6 Date: 2 July 2006

Author's response to reviews: see over
Dear BioMed Central Editorial Team,

Thank you for your answer. I am providing you with a point-by-point response to the reviewer and the revised manuscript. We have made every effort to meet the objections. We would like to comment on the observations (in *italics*) as follows:

**Compulsory revisions**

1) *Methods used for assessing goodness of fit should be described in the methods section of the paper for both the linear and logistic models and preferably the results of those checks should be included in the results section or an appendix.*

Goodness-of-fit methods for both linear and logistic regression have been provided and the Statistical analysis section has been updated.

2) *With respect to the linear regression, the plots for the goodness of fit provided in the author’s response show several problems. The normal probability plot shows that the residuals are clearly non-normal and the plot of the residuals versus the fitted values shows the residuals fanning out which indicates non-constant variance. This means that the linear regression model is not appropriate and the results cannot be trusted. There are many strategies that might help including transformations (have the authors tried all of the plausible Box-Cox transformations?) and the addition of quadratic terms. If none of this helps then the linear model cannot be presented as it is not appropriate and all of the related results and discussion must be removed from the paper.*

Neither the common transformation of the response variable nor the Box-Cox transformation helped in solving the problems of heteroscedasticity. The Goodness Of Fit was improved when excluding the age variable from the independent variables, as shown in the new diagnostic graphs. Residuals are now normally distributed and the heteroscedasticity is significantly reduced.
In addition, univariate, non parametric tests have been added to support the findings. To include all these new findings, the Statistical Analysis section has been updated, the sub-section “Drug use characteristics” in the Results section was rewritten and Table 2 modified. In addition, the picture above has been inserted in the Appendix.

3) The authors have still not provided at any goodness of fit statistics for the logistic models (eg the Hosmer-Lemeshow goodness-of-fit test). The method of le Cessie-van Houwelingen (Biometrics 1991; 47: 1267-82) was applied as described in the Methods section.

1) The authors have still not shown all of the variables included in the model. Age, sex and centre should appear in Table 2 and recruitment unit should be included in Table 4 (since it is implied in the methods section that recruitment unit is included in the logistic models).

All the variables included in the linear regression model are now reported in Table 2.

All the variables included in the logistic regression model are now reported in Table 4. The “duration of drug use” was observed to improve the goodness of fit of the models, so it was included in the logistic regression model.

2) It is still not clear to me how the authors have incorporated “partner’s drug use” into the regressions. Since the setup for this sort of data would typically have a missing value for the response to this question, when a participant does not have a partner most packages would exclude the rows of data relating to participants without partners if the variables were all simply included as is. How have the authors handled this in the analysis?

We included the variable "Partner Status" in the linear regression model. Subjects were categorised as "no steady partner", "steady and married", "steady and not married". Those with "no steady partner" were the reference category (see Table 2).

Regarding the logistic regression model, subjects without a partner were included in the model in the previous version of the paper ("partner's use of heroin" was codified as "no use", "use", and "missing", respectively), and those with missing were not reported in table 4 (because it is almost irrelevant).
However, in the new version of the paper, the diagnostic of the model led us to codify the "partner's drug use" and the "partner status" as a single variable, as now described in the methods and reported in Table 4. All the subject are thus included in the models and are reported in the tables.

3) Why was age dichotomised for the regression analyses? This is not necessary for the regression methods and different choices could result in different outcomes. How did the authors choose the point at which to dichotomise? Originally, age was dichotomised to allow a better comprehension for readers with a poor statistical background, because Odds Ratios on categorical variables are usually much more easily understandable than on continuous variables. To satisfy the reviewer, age is now included as a continuous variable.

4) In the regression analyses were the rows of data for degree level education omitted or combined with secondary schooling? It would be common practice here to combine the highest levels of schooling into one group rather than omitting rows of data because of too few observations. We created a new category: secondary school or higher level of education (see Table 1).

5) Was Fisher's exact test not used for the education variable in Table 1? It would appear necessary for that variable given the small numbers in the degree level category.

Yes, it was. By mistake, we did not use the double star! The ** is now added.

6) Page 7 line -8: The same age cohorts should be compared here since your study contains participants aged from 18 years old and not 15. The description of the results from Figure 1 should be more specific. A higher percentage of your cohort, compared to the national figures presented, appears to have completed middle school and similar percentages have completed vocational school. You need to define what you mean by “lower” educational background.

The description of the results from Figure 1 (Educational status, now page 8) has now been improved in the text. Unfortunately, year by year data on the educational level in the general population were not available and we have been forced to compared our data with the “most similar population” in term of geographical location (Northern Italy) and age (15-49). However, the difference in the distributions is so great that a three-year difference in the cohort is uninfluential.

10) Page 10 Educational Background, last sentence. I do not believe that you have shown this in the paper since you have not asked the participants any questions about their motivation for seeking treatment. How have you reached this conclusion?

The phrase has been changed (it is now on page 11). It should be pointed out that the sentence ”…..seem motivated earlier” was recommended by a previous referee.

7) Page 11 Living and partner status. Your first sentence is too general. You have only shown this for 18-34 year olds. I do not believe that you have shown what you state in the second sentence of this section; you do not appear to have asked the participants any questions about their resources for stability and integration and cannot reasonably infer that family is their sole resource based on the information presented.

The Living and partner status sub-section has been partially rewritten (now on final part of page 12). The Discussion section is designed for hypotheses, interpretations and some conjecture….

8) Page 13 last line. It is not reasonable to say “….a network of support through family and friends” since you have not evaluated whether family and friends are providing a network of support. Simply living with someone does not mean that they are providing support.
Agreed. The phrase “…with significant employment rates and a network of support through family and friends” has been removed.

Minor essential revisions

1) Statistical analysis: the statistic “Chi-squared” not “chi-square”.

Agreed.

2) The authors should explain the direct standardisation method or provide a reference. A reference (n. 13) has been provided.

3) The authors should provide a reference to where they obtain the statistics for all of the general population data. All the information about the general population came from the references n. 9, 11, 12, as declared in Statistical Analyses sub-section (page 6, line 6).

4) The paper would read better if consistent reporting with the number of decimal places was used, certainly for similar things. For example percentages and proportions have been reported with none and 1 decimal place (eg page 6 line -5 should really say 11.2% to match the 88.8% reported on the following line). The p-values reported in the text should be given with the same precision as that reported in the tables. That would help readers pick out the appropriate numbers.

Agreed. We have modified wherever possible all the figures to make the overall text easier to understand, and have harmonised the tables.

5) Page 7 line -9: from Table 1 it appears that only 12% of participants have finished secondary school – where does 24% come from?

24% is the sum of secondary and vocational. We corrected the sentence (in Educational status sub-section, now on page 8).

6) Last sentence page 7: Please reword, I am not sure that I understand what you are trying to say. It appears to me that there is a 14% difference for both these groups.

To improve the comprehension we changed the sentence (in Employment status sub-section, now page 8).

7) Figure 3 and the description in the results (page 8 first paragraph) should match. The Living status sub-section (now on page 9) has been re-written, harmonising the text with Figure 3.

8) Page 8 sentence 5: this is not shown in Table 1.

We agree: it is shown in Figure 3. The error has been modified.

9) Page 8 paragraph on partner status: “Fifty-one percent” not “Fifty”. Also is “slightly more females than males” the right way around?

The Partner status sub-section (now on page 9-10) has been rewritten more clearly.

10) Page 8 line -4. Should 16% not be 33%?

16% refers to male patients with a partner with current drug addiction. The remaining 17% (16% + 17% = 33%) refers to male patients with a partner with drug addiction in the past. The phrase (in History of drug
use of the partner, now on page 10) has been simplified.

11) The acronyms HIV, HBV, HCV and TPHA should all be defined where they first appear in the text and maybe again where they appear in the methods section. Apart from the Abstract, the acronyms appear for the first time in the Background section on page 3. Here they are all defined. As requested, we also modified the acronyms in the Methods section.

12) Table 4. Since the “Partner” variable has been broken down into “married” and presumably “unmarried” steady partners versus no steady partner this should be reflected in the labelling of the categories. Using steady as well as married is confusing since the married group is a subset of the steady group.

Please, see point 6 of Compulsory revision.

13) Page 10 line 1. The prevalence of HIV+ patients from 10 years ago has not been described elsewhere in the manuscript so it is not clear how this is relevant to your results. A statistic on the percentage or prevalence should be given together with a reference. Bibliographic references n. 28 of the previous version of the text, mentions the percentage of HIV 10 years ago. The bibliographic reference has been shown here too (changing the numbers in the bibliography). Now the previous reference under 28 is 16.

14) Page 10 Educational background. Why is it expected that the educational background of participants would be different to the general population? Explain more fully. Female patients only have higher percentages of secondary school completions compared to male patients so this sentence seems quite strong. You have not shown any data comparing males and females educational backgrounds in the general population. Educational background is a protective factor against what? Be more precise in your conclusion here.

“As expected” has been removed. As stated above, The Educational status sub-section has been re-arranged, in answer to the reviewer’s observations.

15) Page 10 line 13. “Substantial” would be a better word than “significant”. Agreed.

16) Page 11 line -9. the authors should be more precise with the percentages here and say 51% and 28% to match the results. Similarly line -4 should say 33% and not 32%. Agreed.

17) Page 12 third last sentence. You need to be clearer about how serological status can affect social integration. It would seem that, if appropriately medically managed, many people in society are integrating well despite such things as HIV status. In what sense do you mean social integration here? The phrase (on page 14) has been re-arranged.

18) Page 12 last sentence. You should quote the national and/or European statistics (and provide references for the national ones). Two new references (n. 31 and 32) have been added. New data have been added to the phrase (now on page 14).

19) A caveat should be added to the “study limitations” section on page 14 regarding the fact that this study contains a non-random sample of patients from a non-random sample of centres. The result of this is that, for example,
comparisons with national statistics may not be reasonable.

A new phrase has been added in the “study limitations” (on page 16).

Other changes

1. The statistical software was changed.