Author's response to reviews

Title: Prevalence and correlates of alcohol and other substance use disorders in young adulthood: A population-based study

Authors:

- Antti Latvala (antti.latvala@thl.fi)
- Annamari Tuulio-Henriksson (annamari.tuulio-henriksson@thl.fi)
- Jonna Perälä (jonna.perala@thl.fi)
- Samuli I Saarni (samuli.saarni@helsinki.fi)
- Terhi Aalto-Setälä (terhi.aalto-setala@hus.fi)
- Hillevi Aro (hillevi.aro@thl.fi)
- Tellervo Korhonen (tellervo.korhonen@helsinki.fi)
- Seppo Koskinen (seppo.koskinen@thl.fi)
- Jouko Lönnqvist (jouko.lonnqvist@thl.fi)
- Jaakko Kaprio (jaakko.kaprio@helsinki.fi)
- Jaana Suvisaari (jaana.suvisaari@thl.fi)

Version: 2 Date: 31 August 2009

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

Thank you for considering our manuscript and for the careful review by the referees. Please find enclosed our revised manuscript “Prevalence and correlates of alcohol and other substance use disorders in young adulthood: A population-based study”, that we now respectfully resubmit for consideration for publication in BMC Psychiatry. We have carefully considered all the suggestions by the reviewers, and revised the manuscript accordingly. However, in our opinion, the suggestion of Reviewer 1 to remove Table 4 and related material, and restrict the paper to presenting prevalence and group comparisons in terms of risk factors, is not well justified. As we argue below, we find that the Reviewer’s view that multivariable analysis used in the manuscript should only be used with longitudinal data with clear causal structure between the variables, is too strict. Based on this argumentation, we decided to retain Table 4 and related material in the manuscript but revise them conforming to the specific suggestions by Reviewer 1. We would also like note that Reviewer 2 did not find the analysis strategy in Table 4 unjustified. In the following, we specifically explain the revisions we have made.

Thank you for your consideration. We look forward to hearing from you.

Helsinki, August 31, 2009

Yours sincerely,

Antti Latvala, MSc
National Institute for Health and Welfare
Dept. of Mental Health and Substance Abuse Services
Mannerheimintie 166, FIN-00271 Helsinki, Finland
Tel. +358 9 19127604
Fax. +358 9 19127570
antti.latvala@thl.fi
REVIEWER 1

Overall statement:
Overall the authors have competently reported prevalence but their attempt to show that behavioral and affective predictors trump most other predictors of substance use disorders is flawed by the cross-sectional design without any indication of time-course even retrospectively, and their interpretations of their findings related to associations needs major revision. After careful consideration, and writing all the points below, I have finally decided that this paper should be restricted to presenting prevalence plus comparison of those with and without SUD in terms of risk factors (Table 3). This means that the whole focus of the paper needs to be shifted and Table 4 and the attendant sections in Results and Discussion should be removed.

Indeed, the whole analysis strategy in Table 4 would be appropriate if the blocks of variables entered had some temporal order and implied causal structure. As the authors say that they are looking only at associations it is not clear why they chose this analytic strategy. After reading this paper I finally decided that this analytic strategy was unjustified, hence my recommendation of the removal of Table 4 and related material.

We thank the Reviewer for her thorough analysis and very careful consideration of the manuscript. However, we consider that her overall judgment of the appropriateness of the analysis strategy is not well justified. The reasons for our disagreement are as follows.

The Reviewer argues that the analysis strategy of studying effects and relative importance of different risk factors (organized in blocks of variables) is not justified because the blocks of variables do not have a clear temporal order and implied causal structure. In addition to this general negative judgment, the Reviewer identifies specific shortcomings in the presentation of the logistic models and the discussion of their results, and gives suggestions as how to complement the analyses and clarify the results in the discussion (see below). We feel that these specific suggestions are well justified and will help to make the results more easily interpretable, and, on the whole, improve the paper. However, in our opinion the Reviewer’s view that multivariable analysis of this kind—assessing the relative importance of different risk factors statistically adjusting for each other’s effect—would only be valid in the case of longitudinal data with clear causal structure between the variables is too strict. The kind of analysis strategy we used in our study is widely used in epidemiological research, in both longitudinal and cross-sectional studies. For example, Modern Epidemiology, 3rd ed., by Rothman, Greenland and Lash (2008), Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd ed., by Cohen et al. (2003), and Applied Logistic Regression, 2nd ed., by Hosmer and Lemeshow (2000) all describe such multivariable models in detail.

In particular, our analyses fit the description of the strategy of examination of alternative hierarchical sequences of independent variable sets by Cohen et al. The motivation for this strategy is given by Cohen et al. as follows (p. 160):

“Sometimes the appropriate sequencing of some variable sets is theoretically ambiguous. Although one usually begins with more distal causes and gradually adds the more proximal causes that may mediate those distal causes, there are times when theory is inadequate to determine such a sequence, when it is likely that there are effects of these IV sets on each other, or when sets are alternative mediators of an unmeasured more distal cause. (--) In such cases the addition of each set to the prediction of Y, over and above the prediction of the other set, would be of interest, and both sequences would usually be reported.”
We think that this description suits well the present situation. For many of the risk factors in our analyses, there is no straightforward temporal and causal sequence. For example, the question on parental alcohol problems was distinctively worded as to cover problems occurring before age 16. Similarly, the direct effect of parental low education is likely to have occurred mainly in childhood and adolescence. But constraining the effects of these risk factors to childhood would be simplistic—and, we think, erroneous—as it is well known that both of these risk factors are influenced by genetic factors and as such are also likely to indicate genetic liabilities whose manifestation cannot be constrained to a certain time period. The same logic applies to many of the present risk factors, e.g. low basic education, learning difficulties, and behavior problems at school. Further, there is evidence of shared genetic factors influencing many of the risk factors—e.g. parental SUDs and behavioral problems in the offspring—which makes a straightforward model with unidirectional causal links between these phenomena impossible.

Relying on this argumentation and taking into account the Reviewer’s critique, we decided to retain the logistic models and related material in the manuscript but revise them conforming to the specific suggestions by the Reviewer. In addition, to emphasize the cross-sectional study design and to tone down the inferences based on the results, we decided to use the term “correlate” in place of the earlier term “risk factor” when describing the variables studied. Overall, we have also slightly edited the text to avoid overly strong inferences considering the nature of the data used.

Logistic models:
The interpretation of Models 1 to 5 (p14-15) fails to acknowledge the problem with the way in which the authors have analysed and presented their results. The leftmost column in Table 4 presents the univariate results from each predictor on at a time. The problem with all subsequent models is that the change in ORs from the univariate model may have occurred because of the relationships within a domain, as well as because of control for age and sex, and whatever other domains have been entered. (--) There are two columns missing which may have been omitted to save space or for which the analyses may not have been done. The first omitted column should have presented the results of entering each block on its own, without including age and sex (heading ‘Blocks’). The column of ‘block’ models would show how much the associations between the variables within each block decreased the ORs. The second omitted column should have been labelled ‘Blocks + age and sex’. This would show whether or not the introduction of age and sex altered the ORs. (--) For Models 1 to 5 there should be a row at the bottom giving either the log likelihood or -2LL. This would enable readers to compare across nested models to see if adding a domain was an improvement, even although ORs for individual variables within the domain may not have been.

Table 4 has been revised as suggested. We now report univariate ORs (the column labeled “Univariate”), adjusted ORs for each variable adjusting for the other variables in that block (“Blocks”), the blocks results further adjusting for the covariates age and sex (“Blocks + age and sex”), and Models I-V as before. For Models I-V we give the log likelihood of the model, and for Models II-V, the likelihood ratio chi square test statistic for the comparison of the model with Model I, and the related p value. We also report these model comparisons on page 15: “Although the AORs for many variables were nonsignificant in Models II-V, these additional domains of correlates clearly improved the statistical prediction of SUD over behavioral and affective factors only, as is evident from the statistically significantly higher maximum likelihood of these models compared to Model I (Table 4).”
**Discussion:**
The discussion of parental factors is marred by a failure to think through possible causal pathways. The two parental factor variables can reasonably be assumed to be temporarily prior to the onset of drinking in the young people. Refer to these factors as A. Refer to behavioural and affective factors as B and substance disorder as C. If the temporal order is A to B to C, then if A and B are entered simultaneously in a regression model to predict C then the coefficient for A indicates only the direct effect of A and not the indirect effect mediated through B. This needs to be clearly discussed.

We have revised the discussion of parental factors on page 19: “Our results suggest that the associations of parental alcohol problems and low basic education with SUD in the offspring are at least partly mediated by the offspring’s attention or behavior problems at school, aggression and anxiousness. This interpretation is based on results from Model II, where the previously significant associations between parental factors and SUD were rendered nonsignificant by behavioral and affective factors in the model.”

**Introduction:**
The authors acknowledge that their data is primarily cross-sectional and that this prevents them from making causal inferences. Consequently it seems rather inflated to claim, as they do in the Introduction, that ‘despite several identified risk factors at individual, interpersonal and societal levels, a detailed understanding of the processes leading to SUDs is lacking’, with the implication that this study will remedy this lack.

This sentence has been replaced by: “Several factors, occurring at the level of individual, interpersonal relations, or society, have been found to increase the risk for SUDs.”

I think that they would be better to mention that most studies focus on one predictor and do not report the relationships between then, instead reporting the relationship between the predictor and the outcome, controlling for possible confounders. They themselves also do not report on relationships between their domains but rather just provide evidence about the impact of including various predictor domains in different models. Therefore revision of the Introduction is required.

The Introduction has been revised as to focus more on independence of different risk factors’ effects of each other, and not on relationships between risk factors. On page 6, we now state that “All in all, the relative importance of different risk factors for SUD and their independence of each other’s effects are not well understood.” We have also revised the Background section of the Abstract, stating: “Several risk factors for alcohol and other substance use disorders (SUDs) have been identified, but it is not well understood whether their associations with SUD are independent of each other.”

**Methods:**
A single five point item on trait anxiousness provides a very limited assessment of ‘affective factors’. There is no way the authors can remedy this problem but they should acknowledge it. Some mention of how comprehensive or otherwise their assessment is in other domains is also warranted.

We have now addressed this issue in the Limitations section of the Discussion (p. 21): “Third, the assessment of the four domains of correlates can in no case be considered comprehensive. For example, the single item used to assess anxiousness arguably provides a very limited assessment of affective factors.”
Minor essential revisions:

P16 Mention that the NCS-R and NESARC were both US studies.

This information is now given.
**REVIEWER 2:**

**Abstract**
The age range of the young adults is not clear. It is not clear if the age range of 21-35 is for the baseline or the follow-up survey.

The age range is for the follow-up. This information is now given in the Methods on page 9.

The summary of results states that “The associations between behavioral and affective factors (attention or behavior problems at school, aggression, anxiousness) and SUD were largely independent of other risk factors,...” However, the findings from Model V of Table 4 do not really support this statement.

This statement is based on the sequence of Models I-V, where the AORs for behavioral and affective factors almost invariably remain statistically significant when adjusted for other factors. In Model V, adjusting for all the variables, the AOR of anxiousness is not statistically significant, whereas the AORs of attention or behavior problems and aggression remain significant even here.

**Background**
Page 5: “Parental SUD also predicts earlier onset of substance dependence in the offspring [23]. Parenting practices, however, moderate the relationships between parental and adolescent alcohol use and problems [24].” The second sentence is not relevant, as there is no measure of parenting practice reported in this paper.

We have removed this sentence and the related reference.

**Methods**
P. 8: In terms of selection bias due to missing data, the authors acknowledge that “Previous analyses indicated that attrition in both study phases was related to age, sex, and education, but not to self-reported mental health disorders or symptoms, including the CAGE scores [39].” However, results in tables 3 and 4 all indicated that missing information for the independent variables is significantly associated with substance abuse/dependence. (--) The authors need to discuss the significant odds ratios for the missing subjects and discuss their implications for the study results.

We discuss the significant odds ratios for missing data in some of the variables in the Limitations section of the Discussion (pp. 21-22): “A fourth limitation has to do with missing data in the Health 2000 variables. In some logistic regression models missingness indicated elevated risk for SUD, reflecting the difficulty of studying individuals with serious substance use problems. However, when all the correlates were assessed, missingness ceased to have a significant effect. In any case, the effect of missing data should be taken into account when assessing the results related to these variables.”

**Measures**
Please clarify, for all variables used in the paper, when the information was obtained, i.e., at baseline or follow-up?

This information is given in Table 1, to which we refer on page 9.
Results
The selection of risk factors and the decisions regarding the sequence in which variables were entered into each analysis are not well justified. It is not clear why the parental factors are considered to be a set of risk factors while age and gender are treated as covariates. It seems that since they are likely to be associated with both substance abuse/dependence and child behavioral & affective factors, these two sets of variables should both be controlled for in all the models.

The logistic models in Table 4 have been revised as suggested by Reviewer 1. The primary focus of the analyses was to assess whether the association between behavioral and affective factors and SUD is independent of the variables from other domains. In the case of parental factors, this question is studied in Model 2.

The discussion of the results in Table 4 should be revised. All the ORs should be AOR (Adjusted Odds Ratio).

The term AOR is now used for all adjusted results.

Table 4 should also be revised. All the ORs should be AOR (Adjusted Odds Ratio).

This revision has been done.

Also, the title for the first column is not accurate, because the results are not from univariate analysis.

The logistic models in Table 4 have been revised as suggested by Reviewer 1. The results given in the column labeled “Univariate” are now from univariate models.

Discussion
P.17: last paragraph: “The findings that early initiation of drinking to intoxication was strongly associated with SUD was anticipated,…….” This statement was not quite true from the results reported in Models III and V in Table 4.

This statement is related to the univariate association between early initiation of drinking and SUD. We have clarified this distinction by organizing the discussion of the logistic models into unadjusted and adjusted associations. In addition, we now also address the adjusted association between drinking initiation and SUD on page 20: “In contrast, the association between early initiation of drinking to intoxication and SUD may be accounted for by smoking and gender, as this association was non-significant when adjusted for these factors.”