In their paper "Urine toxicology, psychiatric symptoms and antisocial personality at intake to treatment predicts abstinence in buprenorphine treatment", Öhlin et al. report on the results of a prospective observational study that investigated to which extent time to involuntary discharge from treatment is predicted by selected baseline characteristics of 123 patients with opiate dependence who participated in an outpatient buprenorphine treatment program with intensive psychosocial therapy elements. The primary predictors were: Number of self-reported symptoms of conduct disorder before age 15 (‘subscale’ of SCID-II screening questionnaire), number of positive test results for different illicit drugs in baseline urine sample (laboratory analysis of urine samples at intake to treatment), severity of self-reported general psychiatric distress (GSI score of the SCL-90-R), and sense of coherence (SOC-29). Observation period ranged from 1 week to 64 months. Cox regression analysis revealed that risk of involuntary discharge during the observation period was statistically significantly associated with higher age, higher number of positive test results for different illicit drugs in urine tests at baseline, higher GSI score, more self-reported symptoms of a of conduct disorder before age 15, as well as less self-reported symptoms of avoidant personality disorder. The authors conclude that patients who have a high degree of poly-substance use, a greater extent of psychiatric distress, and / or a history of conduct disorder may need additional or other types of treatment than an outpatient buprenorphine treatment program with intensive psychosocial therapy elements.

Comments to the authors
General remarks
This is a very interesting paper that has the potential to contribute substantially to the literature on addiction treatment research. It has three major strengths: (1) A prospective long-term follow-up of up to 64 months; (2) an evaluation of a remarkable treatment program for opiate addiction that combines important elements of effective addiction therapy: Treatment duration is long-term, therapy contacts are outpatient, the program is highly-structured and comprehensive with a strict abstinence orientation and the combination of psychological, social and medical therapy elements; (3) outcome data on substance consumption are determined by regular urine analyses. This is outstanding.

However, there are several major revisions to be made. The most important
issues are: 1) In its current state the manuscript contains too many contradictory statements and an often unclear use of terms. 2) Statistical analyses have to be revised. It seems that the authors included simultaneously 19 covariates in one Cox proportional hazard model based on a sample of 123 patients; they have to either cite statistical literature that confirms that the sample size is big enough for including 19 covariates, or they should choose a strategy that is based on stepwise inclusion of covariates to construct a small but valid set of predictors. As a consequence I recommend that this manuscript should be also reviewed by a statistician. The results in Table 2 differ on several points from the results reported in the abstract and text. Taken together there is a strong need for clarification.

Major Compulsory Revisions

• Title: Please clarify: The title of the paper claims that antisocial personality predicts abstinence from illicit drugs in patients who participated in buprenorphine treatment. However, in the results and discussion, you write that the study investigated whether involuntary discharge and dropout from treatment (mainly due to substance consumption) is predicted by the number of self-reported symptoms of a conduct disorder before age 15. The scores derived from the SCID-II screening questionnaire are definitely not sufficient for the diagnosis of an antisocial personality disorder. Does the diagnosis of antisocial personality disorder also predict time to discharge in this study? Concerning the outcome variable it might be more understandable to define a clear primary outcome and to use this term throughout the article, e.g. (1) time to relapse and correspondingly cumulative abstinence probability, or (2) time to involuntary discharge and correspondingly cumulative retention probability.

• Abstract, Methods: "At baseline, subjects were administered the Structured Clinical Interview for DSM-IV-II ..." It should be made clear in the abstract that symptoms of a history of conduct disorder, as well as symptoms of avoidant personality disorder are based on self-report from the SCID-II screening questionnaire.

Abstract, Conclusion: "Patients who have ... more severe symptoms or a history of conduct disorder before intake into treatment may be candidates ..." This statement is misleading. Does it mean 'patients who have a greater extent of psychiatric distress, and / or a history of conduct disorder’? Or does it mean 'patients with more severe symptoms of a conduct disorder or a history of conduct disorder’? "history of conduct disorder before intake into treatment" sounds strange.

• Background, page 3, paragraph 1: "It has long been well-known that a large proportion of patients with addiction relapse during or after treatment [1]. Identifying predictors of risk of relapse in different treatment models may provide valuable information about what type of patients can best be helped in what type of treatment” Please give also some recent references. Is relapse still a major problem in contemporary treatment programs? Is there evidence for any improvements during the last 40 years? Why can identifying predictors of relapse
help to chose better treatment alternatives?

• Background, page 3, paragraph 1: "In an older meta-analysis of predictors of relapse to opiate use, it was found that ..." Were the reported characteristics independent predictors of relapse? Did they differ with regard to predictive power?

• Background, page 4, paragraph 3: "The aim of this study was to study predictors of relapse in a consecutive cohort of buprenorphine treated patients. Based on the literature, we predicted that antisocial personality disorder, polysubstance involvement at baseline, and psychiatric symptoms at baseline, and lower levels on the Sense of Coherence scale would predict relapse." Concerning outcome this is a clear definition: Relapse. Actually it is also time to relapse that is investigated (see also comments to the title). Concerning predictors, the terms should be more specific, e.g.: Number of self-reported symptoms of a conduct disorder before age 15, number of positive test results for different illicit drugs in urine analysis at intake, severity of self-reported general psychiatric distress, and extent of subjective sense of coherence.

• Methods, page 4, paragraph 4, page 5, paragraph 1: Please give information about inclusion and exclusion criteria of the study, and if possible, about inclusion process: How many subjects were eligible? How many patients that were assessed for eligibility did not participate, is there any clinical information about patients who did not participate?

• Methods, Assessments, page 7 and 8, SCID-II: Was there also an assessment of Axis-I disorders with SCID-I? If not, why so? The impact of personality disorder symptoms may be associated with the impact of Axis-I disorders. As can be concluded from the description of subjects on page 9, therapists did also make the SCID-II interview. This information should be given already in the description of assessment. Do SCID-II interview diagnoses of personality disorders also predict time-to-relapse in this study?

There is one major problem with the SCID-II screening instrument: Items concerning features that are not related to conduct disorder before age 15 are easily confused with items concerning current state of psychosocial functioning. The item formulations are simply taken from DSM criteria, and they do not clearly address stable personality features like the items of known personality inventories do. Only the instruction requests that participants rate their experience, feelings, attitudes and behaviors of both their previous and recent life. How did you make sure that patients really judged their personality and not only their current experience of the last months?

The reasons for why only the questionnaire data were used are not convincing: If the questionnaire scores and the interview scores are highly correlated than why not analyzing both data sets? If the interview data are too susceptible to interviewer bias, why are they supposed to be highly correlated to questionnaire data? Are the interview data really not usable when carried out by experienced clinicians? This would mean that thousands of studies based on SCID-II interviews are methodologically not sound. However, there are other reasons to
prefer the questionnaire data: There are good reasons that dimensional measurements of personality disorder traits are more valid than categorical measurements of personality disorders. Perhaps an argumentation that goes in the direction "dimensional versus categorical" is more convincing.

Please give specific results including statistical values of the cited studies to support the statement that the SCID-II screening questionnaire is a valid instrument to measure personality disorder related traits. Please cite also studies with specific results that support the statement that the interview data are not usable because of susceptibility to interviewer bias.

• Methods, Statistical analysis, page 10: There are several issues: (1) Why is the defined primary outcome "relapse" changed to "involuntary discharge"? (2) It seems that the Cox proportional hazard model has 19 simultaneously included covariates. Please do either cite statistical literature that confirms that the sample size is big enough for 19 covariates, or choose a strategy that is based on stepwise inclusion of predictors (as example see Krampe et al. Personality disorder and chronicity of addiction as independent outcome predictors in alcoholism treatment. Psychiatr Serv 2006;57:708-12). (3) The SCID-II screening questionnaire has different numbers of criteria for the different personality disorders. Thus the different "personality disorder subscales" differ with regard to the highest number of points patients can reach, e.g. avoidant personality: 7; narcissistic personality: 16. In order to make the criteria counts of the different personality disorder variables comparable or equivalent, the specific number that the patients reached for a given disorder should be divided by the maximal reachable number of this disorder. In other words, percentages of maximum numbers would be determined. This procedure is similar to the calculation of the subscale scores of the SCL-90-R. (4) Why are the SCL-90-R subscales depression and anxiety mentioned as predictors, what happened to the GSI? According to the description of SCL-90-R and Table 2, only the GSI is used. Why is GSI not mentioned as a predictor but shown in Table 2? This becomes very confusing (see also comments to Table 2).

• Results, page 11, paragraph 2: "The observation period ranged from one week to 64 months. " Can you give the Kaplan–Meier estimate of cumulative abstinence probability at month 64 and mean survival time in days?

• Results, page 11, paragraph 3, Table 2: Results reported in the text and in Table 2 are inconsistent. This is extremely confusing. If I am not mistaken, according to Table 2, the following predictors are not significant: Age, GSI score, number of symptoms of dependent personality disorder. It seems that the impact of number of symptoms of avoidant personality disorder is higher than the impact of number of symptoms of conduct disorder. What happenend to the predictor" Passive-aggressive personality disorder"?

Table 2 is missing 95% confidence intervals of the hazard ratios, as well as regression weights, standard errors of regression weights and degrees of freedom. Should "probability" mean the significance level (p value) of the tests for the different predictors? Names of predictors should be as precise as possible,
e.g. 'number of drugs' instead of 'drugs in urine'. 'Drugs in urine' could also refer to a binary categorical variable.

• Results, Figure 1: The figure caption should be more detailed. In its current state the figure is not self-explanatory. If I am not mistaken, Kaplan-Meier curves estimate cumulative survival probability, not proportions. In the case of this study, the probability may be called cumulative abstinence probability or retention probability; the time-to-event would be accordingly time-to-relapse or time-to-discharge because death is not the event of outcome. Do the three survival curves differ statistically significantly from each other?

• Discussion, page 12: "Even so, the patients with more severe antisocial personality traits were at increased risk of dropping out of treatment." How did you measure severity of traits? If I am not mistaken, the predictor was number of symptoms.

• Discussion, page 12: "Patients with avoidant personality disorder traits ..." This term would refer to the categorical variable "Patients with versus without avoidant personality traits". If I am not mistaken, the predictor was number of symptoms of avoidant personality disorder.

• Discussion, page 13: "In contrast with previous studies, we did not find that sense of coherence predicted substance use outcomes. The reasons for this difference are not clear." Was there also no significant prediction of time-to-discharge when SOC was the only covariate of Cox regression analysis or when it was included in smaller sets of predictor variables? SOC has normally moderate to strong associations with measures of psychiatric distress. Perhaps it would show impact on outcome in a regression model that does not include the GSI as covariate.

• Discussion, page 13: "Self-reported symptoms were independently associated with higher risk of involuntary discharge. Previous research has been mixed concerning the impact of depression and anxiety on involuntary discharge". What does this interpretation mean? Do you refer to SCID screening data, GSI, or the SCL-90-R depression and anxiety subscales? How does the statement fit to Table 2? From which variables were self-reported symptoms independently associated?

• Discussion, page 13: "The use of well-validated instruments to assess personality ...", see comments to Methods, SCID-II screening questionnaire, as well as to statistical analyses.

• Discussion, page 14, Conclusion: "Patients who...have a history of serious conduct disorder." To support this conclusion, the categorical variable "Conduct disorder yes versus no" needs to show a significant impact on outcome. Do the survival curves in Figure 1 differ statistically significantly from each other?

Minor Essential Revisions
• Abstract: "Symptoms Checklist 90 (SCL-90)”: Correct name is 'Symptom Checklist-90-Revised (SCL-90-R)'

• Abstract: "SCL-90 severity, and conduct disorder criteria …” Correct would be 'SCL-90-R general severity score and number of conduct disorder criteria'.

• Abstract: "Criteria for avoidant personality disorder were negatively associated …" correct is 'Number of criteria …'

• Methods, Treatment at the clinic, page 5 and 6: The description of the treatment program is very good. Could you also give information about treatment duration? Could you also explain for which reasons you chose the cut-off of 2 standard deviations above the normative value on AUDIT or SCL-90-R for additional psychiatric interventions? Why not using lower cut-offs, e.g. 1 SD above the norm or deciding on an individual basis?

• Methods, Assessment, page 8, SCL-90-R: "Symptoms Checklist 90 (SCL-90)”: Correct name is 'Symptom Checklist-90-Revised (SCL-90-R)'

• Methods, Assessment, page 8, 9, SCL-90-R, AUDIT: "For the present study, only the Global Severity Index, the mean of all scales, was used …gender-adjusted T-scores … " Which scales are you referring to with "all scales"? Were gender-adjusted T scores used for Cox regression analyses? Why not using raw scores? Gender and age are analyzed as specific predictors in this study, so why should gender-adjusted T-scores be used? The same holds true for the AUDIT? Why using gender and age adjusted T scores when including these variables in the analyses? For the other questionnaire variables, e.g. SOC-29, SCID-II screening questionnaire symptom counts, no adjustment to gender and age was performed. Shouldn't this be consistent?

• Methods, Subjects, page 10, paragraph 1: "The mean AUDIT was 7.98 (range: 0 to 35)". Correct: the mean AUDIT raw score.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests