Reviewer's report

Title: Identification and quantification of change in Australian illicit drug markets

Version: Date: 20 February 2006

Reviewer: Bircan Erbas

Reviewer's report:

This paper uses relatively sophisticated statistical methods to investigate the causes of the 2001 Australia heroin shortage. The methods used are principal components analysis (PCA) and SiZer analysis. PCA has been in widespread use for many years and is helpful for identifying key features in large data sets. SiZer is a new nonparametric smoothing method (developed in the last few years) and is useful for comparing information on different scales.

The methods are applied to 17 time series from NSW, collected over 66 months. Unfortunately, the analysis does not adequately address the research question.

Abstract
The objective of the study isn’t well defined in the abstract. The background section needs to be expanded to state that the aim is to use statistical modelling to investigate the cause of the 2001 heroin shortage and to determine whether it was a natural part of the epidemic cycle or whether it was sparked by independent events outside the drug market.

Introduction
1 The research question could be defined more precisely. I suggest a clear statement of the research question be given at the start of the second paragraph of the “Background” section (p3).

Methods
2 Para 1 of “Methods” section should be titled “Data” or something similar.
3 The data are not adequately described. It is essential that the authors state the time period over which these data were collected.
4 “The data series represent counts . . . ”. Do the counts represent monthly aggregated data? This needs to be clearer in the text.
5 The PCA method is a well-known method and does not need detailed description. However, the authors should provide a reference to a good introductory textbook on the subject (not just the originating paper of 1901). (Jolliffe is mentioned later, but is not really an introductory text.)
6 Page 5, line 2, “. . . particularly relevant for Gaussian data.” But the data in this study are not Gaussian (they are standardized counts). So this statement is irrelevant.
7 There is a problem with the order of material in the discussion of PCA. The multivariate data set needs to be described before the interpretation of loadings (p5).
8 The method of PC rotation (if any) needs to be stated.
9 The SiZer description needs to be explained more carefully as this method is not well-known. For example, what test is used for assessing a significant slope? Is significance at the 5% level? How are end-points handled? Does the test allow for serial correlation in the data? If not, is the test justified for these data?
10 The authors say they are using a “kernel density estimator” whereas they appear to be using a kernel regression estimator where the data are smoothed over time. Please clarify.
11 The statistical package used to conduct the analysis has not been referenced.
Results
12 Figure 1: Lines are very difficult to distinguish in non-colour print. I assume this will be printed in colour.
13 Figure 1: Show the actual calendar months clearly (not month numbers). Show the time of the heroin shortage on the graph. Similar comments for Figures 2, 5 and 6.
14 The “component” of Figure 4 is actually month of observation. It would be much clearer if this was explicit in the graph. That is, plot against calendar months.
15 In Figure 5, how is the “best fit” determined? Being time series data with serial correlation, the usual methods of determining best fit (e.g., cross-validation or plugin bandwidths) will not work here. Similar comments apply to Figure 6.
16 p9. Again, this is described as a “kernel density estimator” when it is a kernel regression estimator.
17 p9. The interpretation of these plots seems “imaginative” to me. What is shown is that there is a significant decrease in heroin overdose deaths around months 48–52, and that this result is independent of the bandwidth used. However it has not been shown that this is a separate phenomenon from the cycle because it has not been shown that the cycle is necessarily smooth. No evidence is presented to show that the cycle cannot include a sudden decrease. Consequently, this decrease may simply be part of the cycle.
18 For the cocaine data, the results of the SiZer plots are never related back to the original research questions.

Discussion
19 I disagree that the PCA analysis is “strong evidence” for the heroin shortage being a separate event from the long-term cycle in the drug market. The authors have not demonstrated anything about the shape and structure of the long-term cycle, so there is nothing to say that the empirical behaviour observed is not part of some longer-term repeating pattern.
20 The lack of correlation between the first two PCs is not evidence against the heroin shortage being part of a natural correction to the drug market because PCs are uncorrelated by design. It couldn’t be any other way.
21 All that the authors have shown is that the most important feature of the observed data occurred around month 50. It is assumed (reasonably) that this is associated with the heroin shortage. The authors have not shown that this feature is a separate phenomenon from the long-term cycle and so the research question has not been answered.