Evaluation of a procedure to assess the adverse effects of illicit drugs

J.G.C. van Amsterdam, # W. Best, A. Opperhuizen and § F.A. de Wolff
Laboratory for Toxicology, Pathology and Genetics, National Institute for Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands

# Coordination point Assessment and Monitoring new drugs, Ministry of Health, Welfare and Sports, The Hague

§ Toxicology Laboratory, Leiden University Medical Center, Leiden, The Netherlands.

Correspondence: Dr. J.G.C. van Amsterdam, RIVM-TOX, P.O. Box 1, 3720 BA Bilthoven, The Netherlands. E-mail: JGC.van.Amsterdam@rivm.nl

Abstract
The assessment procedure of new synthetic illicit drugs that are not documented in the UN treaty on psychotropic drugs was evaluated an using the Electre model. Drugs were evaluated by an expert panel via the open Delphi approach, where the written score was discussed on 16 items, covering medical, health, legal and criminalistic issues of the drugs. After this face-to-face discussion the drugs were scored again.

Taking the assessment of ketamine as an example, it appeared that each expert used its own scale to score, and that policy makers do not score deviant from experts trained in the medical-biological field.

Of the five drugs evaluated by the panel, p-methoxy-metamphetamine (PMMA), gamma-hydroxybutyric acid (GHB) and 4-methylthio-amphetamine (MTA) were assessed as more adverse than ketamine and psilocine and psilocybine-containing mushrooms. Whereas some experts slightly adjusted during the assessment procedure their opinion on ketamine and PMMA, the opinion on mushrooms was not affected by the discussion held between the two scoring rounds.

All experts rank the five drugs in a similar way on the adverse effect scale i.e. concordance scale of the Electre model, indicating unanimity in the expert panel with respect to the risk classification of these abused drugs.

Key words: drugs of abuse, illicit drugs, assessment, Delphi, Electre
Introduction
In 1997 the European Council decided to start an Early Warning Mechanism (EWM) to assess the risk-benefit of new synthetic illicit drugs that are not documented in the UN-treaty on psychotropic drugs. For this reason the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) and Co-ordination point Assessment and Monitoring new drugs (CAM) were erected to support these actions at European and Dutch level, respectively.
The Dutch assessment board (CAM) focuses on all new drugs: new chemical entities of natural and synthetic origin, new combinations and applications of known drugs including pharmaceuticals, mushrooms, herbs, and smart drugs like ecstasy. If applicable within the EWM the new drug will be reported to the European authorities of the EMCDDA.
To assess such new drugs in the described way, the CAM has at her disposal a broadly oriented expert panel that evaluates the various properties of the drug ranging from biological activity to legal and criminalistic aspects. For an overview of the 16 items assessed by the expert panel cf. Table 1.
The assessment procedure comprises of three steps; a written judgement for the 16 items (score 1-5), a discussion about the various judgements using the Delphi approach (Dalkey and Helmer, 1963), and finally a second written judgement.
The aim of the present paper was to evaluate this assessment procedure which retains the character of a multiple criteria approach. The Electre model, as proposed by Roy (Roy, 1990; Roy, 1991), seems suitable to evaluate the assessment procedure used as it represents a multi-criteria analysis based on the concept of outranking.

Methods
Assessment procedure
The assessment procedure consists of three subsequent steps that finally results in a final judgement based on consensus. Consensus is attained by means of the oral Delphi method that, in contrast to the classical Delphi method (Dalkey and Helmer, 1963), is not anonymous but has the same iterative and expert character. The expert panel is represented by members with a different scientific backgrounds (ranging from medical biology and toxicology to criminology) and responsibilities (ranging from policy makers to drug prosecution officers). The number of experts may vary per assessment with the aim to attain a minimum quorum of 10 members.
Several weeks before the assessment procedure is started, the experts receive an expert report (fact sheet), describing the relevant properties of the drug based on open literature and other sources available to the CAM. This information is presented in an objective way (controllable
data) and used by the experts to give their initial written score on a form with 16 items. The 16 items addressed are depicted in table 1, and fall in the following four categories: Individual health, Public health, Violation of civil order, and Criminal involvement. The score ranges from 1 to 5: 1 = no risk; 2 = low risk; 3 = risk is present; 4 = high risk; 5 = very high risk. The written score in the first round is discussed face-to-face by the expert panel, where the individual judgements are presented, clarified, exemplified and explained with the aim to get consensus. In the final round the experts give again their numerical scores from 1 to 5 on the 16 items to consider whether they have adjusted their initial judgement. The final judgement is made by the CAM and is based on the arguments made during the discussions rather than on average of the individual written scores. This procedure has now been evaluated using a modification of the Electre model originally developed by Roy (Roy, 1990; Roy, 1991).

The Electre model

The Electre model is chosen as basis to evaluate the multi-criteria assessment procedure as it enables to weight and quantify the scores given by the different experts. The model is based on the concept of outranking, originally due to Roy (Roy, 1990; Roy, 1991) and may be defined as follows. Option A outranks option B, if there are enough reasons to decide that option A is at least as good as B, while there is no overwhelming reason to refute the statement. Thus outranking is defined fundamentally at the level of pair-wise comparison between a certain drug and drug D being considered.

All drugs are compared with an imaginary reference drug (drug D) for which the score on all criteria is set at 2.99. To each comparison of drug x and drug D, a concordance coefficient $c_{x,D}$ is associated, capturing the idea of ‘at least as good’. There are different ways to define concordance. Here, in the modified Electre model, concordance is defined as the sum of the weights of those criteria for which the score of the drug x is higher than 2.99 (drug D). In addition, the notion of disconcordance ($d_{x,D}$) captures the idea of ‘not considerably worse than drug D’, and is defined as the sum of weights of those criteria for which the score of the drug x is as high as or lower (i.e. $\leq$) than 2.99 (drug D).

In this way the concordance ($c_{x,D}$) and disconcordance ($d_{x,D}$) indices can be calculated separately for each drug per expert or per group of experts (the expert panel). By plotting the $c_{x,D}$-values on the ordinate against $d_{x,D}$-values on the subordinate, the comparison of the different drugs examined is visualised. If necessary one can classify the different drugs, according to imposed limits for concordance and disconcordance, to drugs with higher and lesser adverse effects as compared with the reference drug.
**Results**

The assessment of ketamine is chosen as an example of the assessment procedure. The phencyclidine derivative ketamine (Ketalar®, Ketaject®) is legally clinically used as a surgical anaesthetic, but is also illegally abused due to its hallucinogenic potential. Table 2 denotes the scores for ketamine on 16 items (depicted in Table 1) given by the expert panel in the first round and the second round following the open discussion. As can be referred from table 2, each expert uses its own scale to score so that the average value of the absolute height of scores is irrelevant. There proved to be no difference in scores between experts according to their scientific background i.e. experts trained in the medical-biological field (M-group) do not score deviant from policy makers (non M-group). Following the face-to-face discussion grossly the same total number of adjustments were made by the M-group and non M-group. However, nine adjustments were made by one single member of the M-group, whereas 12 adjustments were made by four out of the five non M-group members. Finally note that more adjustments were made on health aspects as compared to civil order and criminal aspects (15% and 5% of all scores, respectively).

A visualisation of the change in opinion of some experts following the open discussion is presented in Figure 1. Note that a high concordance reflects a much higher risk as compared with drug B, whereas the reverse is true for a high disconcordance. The shift in assessment scores of expert number 1, 4, 5, and 8 in the upper-left direction reflects the lower scores given in the second round i.e. the drug is assessed in the second round as less adverse than before.

Figure 2 represents the assessment scores (total score) of five different drugs given by the individual experts (three drugs are evaluated twice; two drugs only once). Clearly, the drugs p-methoxy-metamphetamine (PMMA), gamma-hydroxybutyric acid (GHB) and 4-methylthioamphetamine (MTA) were assessed as being more adverse than ketamine and paddo’s (psilocine- and psilocybine-containing mushrooms). Whereas some experts slightly changed during the assessment procedure their opinion on ketamine and PMMA, the opinion on mushrooms was not affected by the discussion held between the two scoring rounds. In addition, it is again evident that each expert uses its own scale of scoring. Still, the parallelism in lines nicely illustrates that the experts rank the drugs in the same (or similar) way on the adverse effect scale i.e. concordance scale indicating unanimity in the expert panel considering the classification of these abused drugs.

**Discussion**

This paper shows that the assessment of five drugs by a panel consisting of experts from
different field of expertise and responsibilities leads to an unanimous decision to rank these
drugs with respect to their adverse effects. In addition, the procedure to assess drugs using the
Delphi approach is sound, and suitable to acquire such unanimous verdict.

The adverse characteristics of illicit drugs are manifold and know a large variety of properties.
In addition to adverse health effects like acute and long-term toxicity, psychological and
physical dependence liability, illicit drugs may affect civil order and induce criminal assaults
and violation. To obtain a sound and balanced verdict on the adverse character of drugs 16 items
have been selected to cover these properties. The variety of drug properties evaluated demands
that experts from different disciplines should be represented in the expert panel. Obviously, the
complexity of the assessment i.e. different aspects considered by experts from different
disciplines urges a well-balanced design of assessment. Last but not least, the assessment
procedure may be frustrated by the subjectivity of the experts.

To attain a sound assessment a procedure was selected based on two scoring rounds with an
open Delphi approach in between. It should be noted here, that written information may
introduce bias. This discussion has the aim to mutually exchange basic information. As
expected, the experts evaluated the 16 items in the first round with a wide variation in score. It is
important to note that following the face-to-face discussion, the scores showed a more consistent
view i.e. the variation in scores (expressed as concordance and disconcordance value; cf. Fig. 1)
was diminished. This clearly shows that the discussion is helpful and additive, and pursues some
experts not only to reconsider their view but also to comply more with general view of the panel
as a whole in order to build consensus. In addition, it is surprising to observe that most
adjustments lead to a lower score than given in the first round indicating that the drugs is
considered by the adjusting experts less adverse than previously thought.

With respect to the ranking of five different illicit drugs, it is evident that experts use their
personal scale to give the scores, but finally come to the same ranking order (cf. Fig 2). On the
other hand, the discussion between the two scoring sessions does not dramatically affect the
verdict of the majority of the experts. However, a closer view at the adjustment data of the
ketamine assessment learns that more items were adjusted in the two health categories (15%)
than in the other two categories (5%). Perhaps more importantly, four out of five experts without
medical-biological background adjusted their opinion, whereas only one out of six medical-
biological experts changed his opinion.

The reason of these deviations is not clear. Possibly, the basic information supplied to the
experts in advance to the first round has essentially a medical-biological nature and is not well
understood by the experts that are not familiar with medical-biological terminology (i.e. the non
M-group). As such, these experts profit relatively more from the discussion where information is personally exchanged.

In conclusion, the assessment procedure described gives a reliable judgement about, and enables comparison of the health hazards and psycho-criminal properties of a variety of drugs of abuse.

References

Table 1. Risk criteria used in the procedure to assess the adverse potential of drugs.

<table>
<thead>
<tr>
<th>Individual health</th>
<th>Public health</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Physical dependence</td>
<td>5. Extent and frequency of use</td>
</tr>
<tr>
<td>2. Psychological dependence</td>
<td>6. Vulnerability of the user</td>
</tr>
<tr>
<td>3. Acute toxicity</td>
<td>7. Availability of information on possible effects of the drug</td>
</tr>
<tr>
<td>4. Chronic toxicity</td>
<td>8. Availability of the drug</td>
</tr>
<tr>
<td></td>
<td>9. Reliability of the drug’s pharmaceutical quality</td>
</tr>
<tr>
<td></td>
<td>10. Reliability of the drug’s distribution and sale</td>
</tr>
<tr>
<td></td>
<td>11. Reported nature and extent of incidents</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Violation of civil order</th>
<th>Criminal involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. Annoyance to the general population</td>
<td>15. Criminality with respect to the final product</td>
</tr>
<tr>
<td>13. Increased resort to use violence</td>
<td>16. Criminality with respect to raw products</td>
</tr>
<tr>
<td>14. Impaired reaction time (traffic, labour)</td>
<td></td>
</tr>
</tbody>
</table>


Table 2. Assessment of 16 items (risk criteria) of the drug ketamine by a panel of experts. The number in brackets denotes the score given in the second round if adjusted following the discussion. Only one score is given when the expert did not adjust his/her score. The symbol $\omega$ stands for the arbitrarily attributed weight factor per item. (M) denotes that this expert has medical-biological expertise.

<table>
<thead>
<tr>
<th>Risk criterium assessed</th>
<th>Individual health</th>
<th>Public health</th>
<th>Violation of civil order</th>
<th>Criminal involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>5 6</td>
<td>7 8 9</td>
<td>10 11</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.1 0.1 0.1</td>
<td>0.1 0.1</td>
<td>0.1 0.02</td>
<td>0.02 0.02</td>
</tr>
<tr>
<td>Expert</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 3 3 2</td>
<td>2 3</td>
<td>2 (2.5)</td>
<td>3 (2)</td>
</tr>
<tr>
<td>2</td>
<td>2 3 3.5 1.5</td>
<td>2 3</td>
<td>2.5 2.5</td>
<td>3 2.5</td>
</tr>
<tr>
<td>3</td>
<td>2 3 3 2</td>
<td>2 2</td>
<td>3 2</td>
<td>3 2 (2)</td>
</tr>
<tr>
<td>4</td>
<td>3 (1.5) 4 (2.5)</td>
<td>4 4 (2)</td>
<td>2 2</td>
<td>3.5 (3)</td>
</tr>
<tr>
<td>5</td>
<td>3 2 2 3</td>
<td>3 (2)</td>
<td>3 4</td>
<td>3 3 (2)</td>
</tr>
<tr>
<td>6 (M)</td>
<td>1 1 4 2</td>
<td>2 2</td>
<td>2 2</td>
<td>3 2</td>
</tr>
<tr>
<td>7 (M)</td>
<td>1 1 4 1</td>
<td>2 2</td>
<td>2 2</td>
<td>1 2</td>
</tr>
<tr>
<td>8 (M) 1.5 (1)</td>
<td>2.5 3 (4)</td>
<td>2.5 (2)</td>
<td>2 2</td>
<td>3 (3)</td>
</tr>
<tr>
<td>9 (M)</td>
<td>2 2 3 1.5</td>
<td>2 2</td>
<td>3 2</td>
<td>2 2</td>
</tr>
<tr>
<td>10 (M)</td>
<td>1 3 2 3</td>
<td>2 2</td>
<td>3 2</td>
<td>3 1.5</td>
</tr>
<tr>
<td>11 (M)</td>
<td>1 2 4 1.5</td>
<td>2 3</td>
<td>3 3 5</td>
<td>3 2</td>
</tr>
</tbody>
</table>
Figure 1. Concordance ($c_{x,D}$) and disconcordance ($d_{x,D}$) values representing the assessment score on ketamine calculated from the scores given by 11 experts (cf. Table 2). Open symbols refer to values based on scores given in the first round, whereas closed symbols (and asterisks) refer to values in the second round. Triangles refer to values that were not adjusted by the expert. Expert number 6 to 11 have medical-biological expertise.

Figure 2. Concordance values representing the assessment score on five different drugs calculated from the scores given by 6 to 12 experts. The numbers given as additive to the drug’s name on the ordinate refer to round 1 and 2. PMMA: $p$-methoxy-methamphetamine; MTA: 4-methylthio-amphetamine; GHB: gamma-hydroxybutyric acid.