ELSEVIER

Accident Analysis and Prevention 37 (2005) 843–851

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Drug identification performance on the basis of observable signs and symptoms

David Shinar *, Edna Schechtman

Industrial Engineering and Management, Ben Gurion University of the Negev, Ben Gurion Blvd, Beer Sheva 84105, Israel

Received 16 August 2004; accepted 4 April 2005

Abstract

A double blind study was performed to evaluate the ability of trained police officers to detect drug impairments and to identify the type of drug responsible for the impairment, on the basis of observed symptoms and psychophysical measurements of performance alone. The officers were not allowed to interview the subjects, and their conclusions were based solely on observable signs and symptoms on systematically measured vital signs, and on standardized sobriety tests of motor coordination. Results showed that with this partial information, the officers are able to detect drug impairment at better-than-chance levels with a sensitivity (correct detection of impairments) of 72%, but with a specificity of 43% (false alarm rate of 57%). Furthermore, the association between drug ingestion and identification of the specific impairing drug category was not very high, with sensitivities ranging from a low of 10% for amphetamine to a high of 49% for cannabis. Based on both sensitivity and specificity, drug identification was best for alprazolam impairment, noticeably poorer for cannabis and codeine impairment, and no better than chance for amphetamine impairment. Performance could have been improved if the officers were to list the two most probable impairing drugs (rather than limit their decision to only one), and if they were more consistent in their interpretation of observable signs and symptoms.

Keywords: Forensic science; Drugs and behavior; Drug detection; Drug identification; Drugs and driving

1. Introduction

The use of drugs is a significant traffic safety issue. In a 1996 survey of American households, 4% of the licensed respondents reported that in the past year they have driven within 2 h of ingesting alcohol and drugs, and an additional 1% reported that they have driven after using drugs only (Townsend et al., 1998). To complicate things, drug abuse is most common among the younger male drivers who are already a high-risk group for traffic crashes (Townsend et al., 1998). The detection of impairment from alcohol has received much attention, and an empirical Standardized Field Sobriety Test has been developed and later validated for establishing probable cause that a motorist had a blood alcohol concentration levels of 0.08 and above (Stuster and Burns, 1998).

In contrast to the detection of alcohol above the legal limit, developing a means to determine if an impaired motorist is under the influence of drugs other than alcohol has been much more difficult. Nonetheless, joint efforts by the Los Angeles Police Department, the International Association of the Chiefs of Police, and the National Highway Traffic Safety Administration have yielded a Drug Evaluation and Classification Program (DECP). The program trains police officers to (a) detect whether observed impairment is from alcohol and drugs, and (b) identify the category of drugs responsible for the observed impairment. Drugs are classified into seven categories: CNS stimulants, CNS depressants, narcotic/analgesics, phencyclidine (PCP), cannabis, hallucinogens, and inhalants. The complete evaluation is a 12-step procedure that includes an assessment of the suspects' general appearance, a structured set of examinations of vital signs and symptoms, several tests of pursuit eye movements, the pupils' responses to light, psychophysical tests, and an interview.

* Corresponding author. Tel.: +972 8 647 2215; fax: +972 8 647 2958.
E-mail address: shinar@bgumail.bgu.ac.il (D. Shinar).

0001-4575/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2005.04.002

The ability of the trained officers to detect and identify drugs has been the subject of several field and laboratory validation efforts (Adler and Burns, 1994; Bigelow et al., 1985; Compton, 1986; Heishman et al., 1996). However, all of the previous evaluations suffered from either inadequate methodological controls, or from very limited data. Unfortunately, a completely objective and fully comprehensive evaluation of the trained officers' performance is essentially impossible for several reasons. In the natural environment, and as part of the DECP protocol, the officers interview the suspects and supplement their observations with the suspects' admissions (which, in turn, is often prompted by the partial knowledge the officers gain from their observations). The interview is, therefore, an important component of the evaluation process. In the typical naturalistic setting, the officers have a fairly accurate knowledge of the prevalence of different drugs on the street, and therefore, can adjust their evaluations based on their estimates of prior probabilities for different drugs. In the natural environment, the suspects are typically arrested for suspected impaired driving, and thus have already manifested obvious impaired behavior. Finally, the potency of abused drugs typically exceeds the therapeutic prescription levels (Preston et al., 1992) and may be much more potent than safe doses used in a controlled administration study. Given these limitations, the purpose of the present evaluation was not to provide a comprehensive evaluation of the officers' performance, but rather to evaluate the benefits – and potential role – that the physical signs and symptoms can provide towards the detection and identification of drug impairment with (relatively) safe doses, by specially trained police officers, without prior selection on the basis of observable impairments, without the benefit of an interview, and without knowledge of prior probabilities of different drugs.

This study evaluated the ability to detect drug impairment and identify the impairing drug category, solely on the basis of the tests and observable signs and symptoms. Four drugs – corresponding to four different drug categories – were evaluated in this study: (a) cannabis, represented by marijuana cigarettes; (b) depressant, represented by alprazolam; (c) narcotic analgesic, represented by codeine; and (d) stimulant, represented by amphetamine. The evaluations were assessed primarily by measuring the sensitivity (true positives) and specificity (correct rejections) levels.

A previous analysis of the officers' performance on the same subjects was conducted and reported by Heishman et al. (1998). However, their analysis relied on different interpretations of the officers' written report. In their report, the officers were required to note all observable signs and symptoms and then, state their conclusion regarding presence of impairment. If impairment were noted, then they were required to name the source of impairment in terms of one or more of seven drug categories. Unfortunately, in approximately half the cases, the officers checked the 'unimpaired' category, and at the same time cited one or more sources of drug impairments. This is not a practice recommended by the DECP procedures, and it can only be surmised here that it reflects a lower level of confidence concerning impairment for driving. Heishman et al. (1998) dealt with this discrepancy by ignoring the drugs cited by the officers whenever the 'unimpaired' category was checked. This created a large subset of 'unimpaired' decisions. In our analysis, we relied on all of the officers' reports of drug impairment (even when the officers checked the 'unimpaired' category). This approach yielded a much smaller set of unimpaired cases, and is arguably a better indicator of the officers' sensitivity to drug impairments.

As noted, this validation study was particularly severe in its criteria relative to the setting in which officers' typically detect and identify drug impairment. Therefore, it is not an evaluation of the DECP program and the officers' skills at using the program, but only an evaluation of the observed signs and symptoms and predictors of drug impairment. Still, it is important to see to what extent the physical signs and symptoms alone account for specific drug impairments, and how well do the officers utilize the information obtained from the physical signs and symptoms.

2. Methods

2.1. Subjects

The subjects in this study were paid volunteers, recruited by the Addiction Research Center from the Baltimore, MD area through newspaper advertisements. All the subjects were self-admitted regular users of at least one of the study drugs (heroin, morphine, or prescription opiate for the study drug of codeine; marijuana and hashish for the study drug of cannabis; any benzodiazepine, barbiturates, or alcohol for the study drug of alprazolam; and cocaine for the study drug of amphetamine). The subjects had to assure the staff that they never had, neither did they currently intend to seek substance-abuse treatment. They were judged to be in good health, based on a blood analysis and a routine examination of the vital signs. All subjects provided a written informed consent according to guidelines for the protection of human subjects of the Department of Health and Human Services.

2.2. Procedure

A detailed description of the experimental procedure and collection and analysis of blood samples is provided by Heishman et al. (1998). In brief, each subject was individually tested during the regular weekday working hours. All sessions started in the morning. At the beginning of each session, the subject was given a medical examination of his/her vital signs, an evaluation for psychophysical impairments (consisting of cognitive performance tests and tests of oculo-motor control), and blood and urine specimens were taken to check for the presence of drugs and drug metabolites. The blood was sent to a lab off the premises, and the urine specimen was screened on the premises for amphetamines, barbiturates,

benzodiazepines, cocaine metabolites, methadone, phencyclidine, and cannabinoids. After the tests' completion, the subject was given one more examination of vital signs and if all signs were normal, then the subject was dosed by a nurse under a doctor's supervision. The dosing procedure was standard for all conditions. In each session, each subject received two capsules and two cigarettes, but except for one active drug (belonging to one of the drug categories the subject reported using), all the rest were placebo. To allow for the different absorption rates of the different drugs, the dosing and testing procedures were as follows: initial dosing began with 1 capsule of amphetamine or placebo. A second capsule of alprazolam or codeine or placebo was given an hour after initial dosing. Two cigarettes were given 110 min after initial dosing. The officer's evaluation was begun 140 min after initial dosing.

With this time course, the evaluation began either 140 min after amphetamine dosing, or 80 min after codeine or alprazolam dosing, or 30 min after cannabis dosing. The officer's evaluation lasted 20–30 min.

The officers were told that the subjects might be under the influence of none, one, or a combination of two or more drugs (though in reality each subject was given only one drug or a placebo). They were further told that the drugs could belong to any of five categories they were trained to identify – CNS depressants, CNS stimulants, PCP, narcotic analgesics, and cannabis – or any combination of these five drugs, or no drugs at all.

2.3. The drugs evaluation procedure

The evaluation of drug impairment was an abridged form of the standard procedure used in the Drug Evaluation and Classification Program (DECP), since it did not contain any interviews with the subjects. This, of course, makes the officer's task much more difficult, and consequently, the results here should not be interpreted as a true reflection of the total performance capability.

A detailed description of the DECP tests is provided elsewhere (Kosnoski et al., 1998; NHTSA, 1991; Smith et al., 2002). The evaluation consisted of 10 major components given in the following order:

(1) Breath alcohol test, given to measure blood alcohol concentration (BAC).
(2) Preliminary examination, consisting of the observations of general physical appearance, use of corrective lenses, and appearance of the eyes.
(3) Examination of the eyes, including tests for horizontal and vertical gaze nystagmus and lack of convergence.
(4) Divided attention psychophysical tests. Four tests that measure motor control, balance, and the ability to count and estimate time. The tests include (a) walk-and-turn test, in which the subject has to walk heel-to-toe along a line, 9 steps in each direction; (b) one-leg-stand test, in which the subject has to stand on one leg, keep both arms at the sides of the body, and count for 30 s. The process is then repeated while standing on the other leg; (c) Romberg balance test, in which the subject has to stand straight with the eyes closed for an estimated 30 s; and (d) finger-to-nose test, in which the subject has to touch the nose with a finger six times while the eyes are closed and the two hands are outstretched to the sides.
(5) Vital signs examination. Tests of pulse rate, blood pressure, and temperature.
(6) Dark room examination. Measurements of the pupil size under different light conditions (in normal room light, in darkness, and with direct light into the eye) and the reaction of the pupils to light.
(7) Examination of muscle rigidity.
(8) Examination of skin surface for injection sites.
(9) Opinion of the officers as to whether or not the subject was drug-impaired, and if so, what were the likely drug categories causing the impairment. The officers' protocol allowed them to list two categories, and the conclusion was considered valid if one of them were identified in the toxicological evaluation.
(10) Toxicological evaluation. The samples were collected several times in the course of the evaluation and sent for a toxicological evaluation.

2.4. Study design

Each of the subjects was recruited as an ARC in-patient for a period of up to 3 weeks. Within that period, each subject was tested on six sessions, separated from each other by at least 48 h. Prior to the first session, as part of the 'in-take' procedure, the subject's tolerance to the high dose level of his/her drug category was tested by administering the drug at the high dose level. For the actual testing each subject was orally administered each of the three drug levels (placebo, low dose, and high dose) twice. The order of the dose levels was counterbalanced between subjects.

The two dose levels for each drug were:

- Codeine: 60 and 120 mg (in capsules).
- Alprazolam: 1 and 2 mg (in capsules).
- D-Amphetamine sulphate: 12.5 and 25 mg (in capsules).
- Marijuana cigarettes: two machine-rolled cigarettes, each weighing approximately 900 mg. In the low-dose condition (16 mg THC), one cigarette contained 3.58% active THC and one was a placebo cigarette and in the high-dose condition (32 mg THC), each of the two cigarettes contained 3.58% active THC.

The dose levels of the orally administered drugs were selected so as not to exceed the maximum daily therapeutic dose, and so that the effects at these levels should disappear within 8 h after drug administration. The levels used were consistent with the range of levels used in other studies by Evans et al. (1994) and Kelly et al. (1997) for alprazolam; Sutton (1983) and Perez-Reyes et al. (1988) for marijuana; Linnoila and Hakkinen (1974) for codeine; and Perez-Reyes

et al. (1992) and Tedeschi et al. (1983) for amphetamine. In the case of marijuana, the originally planned dose levels were reduced by half after two subjects in a previous study showed adverse responses to the high-dose levels.

The study design called for the participation of 12 subjects in each of the drug categories, making a total of 48. However, not all subjects who began the protocol completed it, and of those who did, not all of the subjects completed all of the evaluations. Consequently, in order to maximize the size of the data set, this analysis included all of the data available from 54 subjects who were entered into the study. The total number of sessions available was 302, but two sessions in which the officers identified PCP as the impairing drug were omitted from the analysis. Although the complete set of six evaluations was completed only for the amphetamine-dosed group, the total number of sessions with each drug and each level of each drug was approximately the same, varying from a low of 23 to a high of 28. A total of 39 officers participated in the study. The rotation of the officers did not match the schedule of the subjects and consequently, each subject was typically tested by more than one officer and each officer evaluated more than one subject.

3. Results and discussion

The analyses presented below focus on two issues. The first issue is an evaluation of how well drugs can be detected and classified by trained officers on the basis of observable signs and symptoms only. The second issue is which specific signs and symptoms the officers tend to rely on, and how useful are they in reaching the correct conclusions. Thus, the first set of analyses was aimed at evaluating the validity of the officers' conclusions, whereas the second set of analyses was aimed at evaluating the methods the officers used to reach their conclusions.

3.1. The validity of the officers' conclusions based on signs and symptoms only

To evaluate the detection of impairment and diagnosis of the drug category, the officers' conclusions were cross-tabulated with the actual drug dosing. These tabulations were then analyzed to determine the existence of associations between the officers' decisions and the actual dosings, and if so, what are the strengths of the associations between the two. For the determination of the existence of an association, the standard procedure is chi-square analysis. To determine the strength of the relationship, different measures had to be used, since no measure was adequate for all situations. In general, for 2×2 tables, phi coefficient of correlation and the contingency coefficient were used. For larger tables, Cramer's V measure of association was used. To determine actual levels of agreement, kappa coefficient was used. The kappa coefficient measures the accuracy of the officers' conclusions in terms of the matches between the drugs given and the drug categories assigned, rather than just the degree of non-random associations between the two. The final measure used was the uncertainty coefficient (Uc)—a measure based on the information metric that indicates the amount of relative uncertainty that is reduced about the drug that was administered by knowing the officers' response [Uc (drug|officers)]. When the officer's conclusion reduces all uncertainty, i.e., when it provides perfect identification, Uc = 1.0. When the officer's conclusion essentially reduces no uncertainty, i.e., it does not provide any valid information, Uc = 0.0. The benefit of the Uc is that it provides information on the added value of the officer's conclusions relative to simply using the a priori probability about the drug's prevalence in the sample of subjects based on the study design.

For specific levels of performance, we also calculated sensitivity, specificity, false alarms, and misses. The letters are defined in Table 1.

$$\text{Sensitivity} = \frac{d}{c + d},$$

the probability that the officers will detect the specific drug impairment.

$$\text{Specificity} = \frac{a}{a + b},$$

the probability that the officers will identify non-drug-impaired subjects as such.

Complements to the measures of sensitivity and specificity are misses and false alarms, respectively, which are commonly used in signal detection theory (Green and Swets, 1966).

In evaluating the officers' performance, identical analyses were conducted on two data sets: one that included the placebo, the low-dose and high-dose conditions of each drug, and one that included only the placebo and high-dose conditions of each drug. The inclusion of the low-dose cases provided a larger *N* of dosed subjects for each drug (increasing the power of the tests). The exclusion of these cases presumably reduced the variance in the signs and symptoms due to drug impairments and presumably provided the officers with a more clear-cut distinction between the drug-dosed and the placebo-dosed subjects. Interestingly, the results of the statistical analyses were similar in both cases, and therefore, we only report the analysis that included both the high- and low-dose levels. One possible explanation for the similarity is that relative to abuse levels, both the 'low'-

Table 1
A generic representation of the classification of subjects into dosed vs. not dosed as a function of their actual condition

Actual condition	Predicted condition		
	Not dosed	Dosed	Total
Placebo/normal	a	b	a + b
Dosed	c	d	c + d
Total	a + c	b + d	a + b + c + d

Table 2
Actual dosing (including high- and low-dose levels) vs. officers' evaluation of impairment

Actual dosing	Officers' evaluation		
	Unimpaired	Impaired	Total
Placebo	**44 (43.1)**	58 (56.9)	102 (100)
Drug	55 (27.8)	**143 (72.2)**	198 (100)
Total	99 (33.0)	201 (67.0)	300 (100)

Numbers in parentheses indicate percent evaluated relative to actual.

and 'high'-dose levels – limited by ethical constraints – are actually low and quite similar.

3.1.1. *Officers' ability to detect drug impairment*

The officers' ability to recognize drug impairment – with both low and high doses – relative to the placebo condition is given in Table 2, where the correct decisions – the diagonal entries – are in bold.

Fisher's exact test revealed that the officers' ability to detect drug impairment was significantly better than chance ($p = .009$). However, the measures of association which describe how well, or how much better than chance the officers perform in detecting drug impairment were disappointing; with a phi correlation of 0.16 (with a 95% confidence interval (0.04–0.27)), a contingency coefficient of 0.15, and a kappa coefficient of agreement of 0.16. These values are quite low relative to the perfect level of association of 1.0. In particular, the uncertainty coefficient indicates that the amount of uncertainty reduced by considering officer's decision was almost nil: 0.02 out of a maximum of 1.0 (with a 95% confidence interval (−0.01–0.05)).

The analysis of the officers' sensitivity and specificity showed that the sensitivity (the detection of impairment given drug dosing) was moderate at 72%. The specificity (the ability to assess unimpairment in the placebo condition) was near chance at 43%. The complementary miss rate was 28% and the complementary false alarm rate was 57%. This high false alarm rate may be due to two reasons. First, the officers' false assumption that most subjects had ingested drugs. Second, the fact that whenever an impairing drug was noted – even when no conclusion of impairment was listed – it was listed as a positive decision of drug detection. In summary, the different analyses were all consistent in showing that the officers' ability to detect drug ingestion and differentiate between drug impaired and unimpaired subjects, although better than chance, was quite poor.

3.1.2. *Identification of specific impairments relative to the drugs ingested*

The officers' ability to identify the specific psychoactive drug classes was assessed by cross-tabulating the officers' principal conclusion relative to the actual dosing. To be considered as a correct identification, the DECP protocol allows the officers to list up to two drug categories (out of the possible seven) when they suspect drug impairment. Still, to be conservative, for this analysis only the first category – which reflects the officer's opinion of the most likely drug that is impairing the individual – was included. Again, the analysis included both low-dose and high-dose cases. It is important to note that the officers were told in advance that hallucinogens and inhalants are not included in the study.

The results of this analysis are summarized in Table 3. If the officers' performance were perfect, the only non-zero entries would be in the diagonal cells. Although this is not the case, performance is definitely much better than chance, as supported by a chi square = 72.27, $p < .001$. Identification appears to be best for the depressant alprazolam and placebo (i.e., lack of impairment) and worst for the amphetamine stimulant.

Two summary measures of association are appropriate for this table: the contingency coefficient, which was 0.44 and the kappa coefficient of agreement, which was 0.152. The discrepancy between these two measures shows that although specific drug impairments are distinguishable from each other (chi square and contingency coefficient), the officers' levels of accuracy – the actual attribution of impairment to the correct drug category (kappa coefficient) – is quite poor. The uncertainty coefficient was 0.08, reflecting a very small added value of the officers' conclusions. However, what these statistics fail to reveal is that when the officers err, the error is not a random one. The value of Table 3 is that it reveals the specific confusion errors that officers consistently make:

(1) Cannabis was most often correctly recognized (yielding relatively high sensitivity), but at the cost of relatively low specificity. Specificity was low because it was often con-

Table 3
Drug administered (low and high doses combined) vs. first drug category listed by officers

Drug dosing	Officers' decision					Total
	Cannabis	Depressant	Narcotic	Stimulant	Unimpaired	
Marijuana (cannabis)	**15 (30.6)**	8 (16.3)	10 (20.4)	4 (8.2)	12 (24.5)	49 (100)
Alprazolam (depressant)	4 (8.2)	**21 (42.9)**	18 (36.7)	1 (2.0)	5 (10.2)	49 (100)
Codeine (narcotic)	7 (14.3)	6 (12.2)	**18 (36.7)**	1 (2.0)	17 (34.7)	49 (100)
Amphetamine (stimulant)	20 (39.2)	4 (7.8)	2 (3.9)	**4 (7.8)**	21 (41.2)	51 (100)
Placebo (unimpaired)	21 (20.6)	20 (19.6)	15 (14.7)	2 (2.0)	**44 (43.1)**	102 (100)
Total	67 (22.3)	59 (19.7)	63 (21.0)	12 (4.0)	99 (33.0)	300 (100)

Numbers in parentheses represent percent identified relative to actual administration.

fused with either no impairment, narcotic-based impairment, or depressant-based impairment.

(2) Depressants are just as often mistaken for narcotic analgesics as they are recognized as depressants. However, they are not confused with other drug categories or with placebo. Analysis of the high-dose condition only (not shown here) showed that the confusion with narcotics is slightly reduced and the confusion with unimpairment is greatly reduced when the low-dose cases are eliminated.
(3) Narcotic-analgesics are most – and equally – likely to be either correctly identified or totally missed (in which case the subjects are judged as unimpaired). When the low-dose cases are eliminated, the officers are slightly less likely to judge the high-dosed subjects as unimpaired.
(4) Stimulants, at least D-amphetamines at the doses administered here, are the most difficult to identify. Subjects under its effects are more likely to be erroneously classified as either cannabis-impaired, or not impaired at all. For this drug, elimination of the low-dose cases did not improve the officers' performance levels.
(5) Placebo-dosed subjects are more likely to be classified as unimpaired than impaired by any other specific drug. However they are nearly half as likely to be misclassified as impaired from cannabis (21%) or from a depressant (20%).

The officers' confusions among drug categories were also apparent from their choice of a second category of impairment. In approximately 50% (107 cases) of the 201 cases where the officers noted impairment, they also selected a second category. The most common joint citations were cannabis and narcotic-analgesics (31/94 = 33%) and cannabis and depressant (28/94 = 30%). These, in fact, were the drugs that were most commonly confused with actual cannabis impairment. Also, probably because of their biased past exposure, the officers cited cannabis – either as a primary or secondary drug category – in half of all cases where they noted impairment (102 out of 201 cases). Of all the abused drugs, cannabis is typically the most prevalent drug in drivers suspected of impairment (Ward and Dye, 1999).

To assess the officers' ability to correctly identify each of the specific drug categories, independently of the others, we analyzed the data using 2×2 tables of the actual dosing versus the officers' decision. An officer's decision was coded as correct if the drug administered belonged to either the first or second drug category identified by him (as allowed by the DECP protocol). An incorrect decision was either one of unimpairment or impairment due to other drug categories. The summary statistics for each of the drugs are presented in Table 4. Although the phi correlation, contingency coefficient, and kappa coefficient measure different characteristics of association, the numerical values were quite similar, and therefore, only the phi correlations – the most familiar of these statistics – and the uncertainty coefficient are reproduced in Table 4.

The obvious finding is that the ability to determine the drug used on the basis of observable signs and symptoms, was quite variable. Given the high rate of false alarms with therapeutic dose levels, the ability of the officers to detect drugs in all categories was very low. Still, comparisons among the drug categories indicate that performance is best with alprazolam, worse with cannabis and codeine, and worst – being no better than chance – with amphetamine. In the case of amphetamine, the high specificity (91%) simply reflects the tendency to avoid citing that category (possibly because it is rarely encountered on the street) unless its signs are extremely patent. Amphetamine dosing (as seen in Table 3) was most often either mistaken for cannabis impairment or missed altogether. Perhaps the most revealing and discouraging are the uncertainty coefficients, which were either low or actually zero (in the case of amphetamine).

3.2. Officers' reliance on specific signs and symptoms

To provide some insights into the officers' decision processes, an analysis was done to see which specific signs and symptoms they used to identify the different drug categories. Analysis of variance (ANOVAs) tests were conducted in which the officers' conclusions – in terms of the four drug categories (plus 'unimpaired' conclusion) – was treated as the classifying variable, and the performance measures on the specific signs and symptoms were treated as the dependent variables. With Dunnett's post-hoc test procedure (treatments versus control), we then examined whether or not the mean performance score of each sign or symptom for each of the drug categories was significantly different from the mean score in the 'unimpaired' judgement condition. These analyses were conducted on all the signs and symptoms for which

Table 4
Comparative evaluations of the officers' performance in correctly identifying the four different drugs (using both high-dose and low-dose drug administrations, and both first and second categories of impairment)

Drug	Chi square	Sensitivity (%)	Specificity (%)	False alarms (%)	Misses (%)	Phi correlation	Uncertainty coefficient
Cannabis	5.86*	49	69	31	51	0.14 (0.02–0.26)	0.02
Alprazolam	16.24**	47	80	20	53	0.23 (0.11–0.36)	0.05
Codeine	5.58*	45	72	28	55	0.14 (0.02–0.26)	0.02
Amphetamine	0.02	10	91	9	90	0.01 (−0.11–0.12)	0.00

* $p < 0.05$.
** $p < 0.001$.

interval-scaled measures could be derived; namely nystagmus, pupil diameter under the different light conditions, pulse and blood pressure, temperature, and the sobriety tests of one-leg-stand, Romberg balance test, walk-and-turn, and finger-to-nose test. The results of these analyses showed that:

(1) The officers relied on all four psychophysical tests and horizontal gaze nystagmus to conclude that a person is impaired, regardless of the selected impairing drug category. This was deduced from the fact that the average performance scores on the nystagmus test and on all of the psychophysical tests were significantly poorer whenever any impairment was identified than when the officers concluded the subject was not impaired. This reliance was not always appropriate. For example the DECP guidelines indicate that nystagmus is characteristic only of depressants (e.g. alcohol, alprazolam) but not of the other three categories—narcotics, stimulants, and cannabis. Yet the officers occasionally noted nystagmus and still concluded that the impairment was due to one of these latter categories. In this respect, the officers often reached conclusions that were inconsistent with the DECP.
(2) In making a specific diagnosis of the impairing drug category, the officers apparently used only one or two 'pivotal' signs/symptoms to guide them in their decision. Thus, in addition to the psychophysical tests and nystagmus, the officers typically noted only one measure that was significantly different from their 'unimpaired' judgment. For their identification of cannabis as the impairing drug, the officers noted a raised pulse rate. For identification of a depressant, they relied on a raised temperature (and possibly reduced pupil diameter under direct light). When they believed the impairment was due to a narcotic/analgesic, it was based on a lower temperature and a slightly constricted pupil under direct light. When they believed the impairment was due to a stimulant, they relied on an enlarged pupil in the dark and an increase in horizontal gaze nystagmus. Although this approach simplifies the officer's task, it is not sensitive enough to the true complexities of drug effects, and consequently, it is also likely to lead to erroneous conclusions.

In summary, it appears that the officers tended to base their diagnosis primarily on one or two signs or symptoms, and then ignore the remaining signs and symptoms even if they were inconsistent with the DECP recommended guidelines for identification of that drug impairment. This reinforces the conclusion that the officers have a difficulty in simultaneously evaluating all of the information available in all of the observed signs and symptoms.

4. Conclusions

This study focused on the validity of the conclusions that were reached by the officers or could have been potentially reached on the basis of manifest signs and symptoms without the benefits of interviewing the subjects, without having the subjects screened based on initially observed impairment, without street-level drug doses, and without knowledge of the a priori probabilities for the presence of the different drugs (which were definitely lower than the a priori probabilities in the officers' real-life work).

Also, it is important to note that the officers' performance is obviously limited by the data they have been trained to collect and the interpretation of that data as provided by the DECP. Thus, there is a ceiling level of performance that the DECP enables. In a separate analysis (Shinar and Schechtman, 1998), it was shown that the DECP – with the data collected by these same officers – correlates moderately with actual impairments from cannabis, alprazolam, and amphetamine (with phi correlations of 0.57–0.62), but correlates poorly with codeine impairment (phi = 0.13).

With these limitations in mind, the following conclusions emerge from the preceding analyses:

(1) Officers are able to detect drug impairment at statistically significant levels above chance, but have high false alarm rates. This is true for therapeutically safe levels, which are typically lower than levels in drug abuse. It is possible that detection levels would be significantly higher (and false alarm rates lower) with higher dose levels.
(2) The identification levels of specific impairing drug categories are generally low and quite variable: best for the depressant alprazolam, lower for cannabis and codeine, and no better than chance for amphetamine.
(3) Based on their first choice of impairing drugs, the officers correctly identified cannabis impairment in 31% of cannabis-impaired subjects; correctly identified depressant-based impairment in 43% of the alprazolam-impaired subjects and correctly identified narcotic impairment in 37% of the codeine-impaired subjects. For these three drugs, the officers were more likely to identify the drug category correctly than to confuse it with any other specific category. The officers performed much worse with amphetamine impairment, identifying only 8% of the subjects as stimulant-impaired. When the subjects were not drugged (placebo), the officers correctly identified them as unimpaired in 43% of the cases (Table 3).
(4) When the officers determined that the subject is impaired, they listed only one drug category in 47% of the cases. They did this despite the fact that they were allowed to list two categories, and thus increase their accuracy, instead of unwittingly reduce it. When the officers named two possible categories of drug impairment, their correct identifications improved to 49% for cannabis, 47% for alprazolam, 45% for codeine, and 10% for amphetamine (Table 4 versus Table 3).
(5) The officers' identification errors were not random but systematic, and these confusions were consistent with DECP common signs and symptoms. Thus, when two

drug categories were allowed, cannabis was confused in approximately 20% of the cases with a depressant and in approximately 20% of the cases with a narcotic. Alprazolam was more likely to be confused with a narcotic (approximately 50%) than to be correctly identified as a depressant. Codeine was more likely to be correctly identified (45%) than to be missed (approximately 30%). Amphetamine was most likely to be either totally missed (approximately 45%) or identified as cannabis (approximately 35%).

(6) The officers' ability to identify amphetamine-dosed subjects was actually worse than chance (indicative of a specific bias towards confusions with cannabis or lack of impairment). This was due mostly to "misses". The officers rarely named a stimulant as the impairing drug (9% of all impairments: 12 times out of 201 first category choices, and 16 times out of 94 s category choices).

(7) The officers relied on the psychophysical tests and horizontal gaze nystagmus almost exclusively to detect impairment in general, rather than to rely on it, in order to distinguish among drug categories. With respect to the psychophysical tests, this approach is consistent with the DECP procedures. However, with respect to nystagmus, the DECP guidelines indicate that it is symptomatic only of impairments due to depressants (like alcohol), but not to other drug impairments.

(8) In order to identify specific drug impairments, the officers tended to rely on one or two specific 'pivotal' symptoms. The pivotal symptoms were all consistent with the DECP manual. For cannabis decisions, they primarily relied on raised pulse rate. For depressant decisions, they primarily relied on low blood pressure. For narcotic decisions, they primarily relied on lower body temperature and constricted pupil (but only under direct light). For stimulant decisions (which were rarely made), they primarily relied on an enlarged pupil under direct light (i.e., little constriction in response to light). Unfortunately, this approach limited their ability to utilize the wealth of information they had observed and recorded. They were, therefore, under-performing relative to the level that could have been expected had they used their own data and the DECP guidelines consistently.

5. Recommendations

Two types of recommendations stem from this study. The first applies to improvements possible within the present DECP guidelines, and the second involves a thorough re-evaluation of the DECP itself. Within the current DECP guidelines, the simplest recommendation is that when the evidence available to the officers indicates poly-drug use, they should list two drug categories. This is important because it is a legitimate way of increasing probability of correct identification. Second, until a better procedure is developed to identify drug impairment on the basis of observable signs and symptoms, the DECP training procedures should also put more emphasis on analyzing combinations of signs and symptoms rather than tending to rely on a single conspicuous sign. Although the training examples already do that, it appears that this is not sufficiently instilled. In a separate study of the DECP program made by the same authors (Shinar and Schechtman, 1998; Schechtman and Shinar, 2005), we recommend that the officers' training should include the use of formal models developed in that study. These models provide weights to different signs and symptoms in the context of impairments from different drugs.

Also, based on our findings, we propose two different research programs. First, within the current DECP, a follow-up evaluation study should be attempted in which the officers would be required to first evaluate the physical signs and symptoms and then, on the basis of their findings, make an evaluation about impairment. Then they should study the arresting officer's report and interview the subject and on the basis of the additional information be given a chance to revise their conclusions. This method would provide a means of (a) field validation of the results obtained in this study and (b) quantifying the added value of the officer's report and interview information.

Second, we recommend that a series of investigations be conducted on each of the drugs or drug categories that may be of interest in the context of impaired driving, or psychomotor impairment in general (which is then relevant to the workplace). The current DECP, which has remained essentially unchanged since it was initiated approximately 30 years ago, suffers from a dearth of scientific data. It is, therefore, recommended that a series of programs be initiated that would (a) review the current literature relating to the dose response of each drug and each relevant behavior, (b) investigate drug effects on relevant behaviors that have not been previously studied, develop a drug assessment test that would be practical for use by police, and (c) validate the value of the test as practiced by trained officers in a controlled study. A first step in this direction has been taken by Jones et al. (2003) who conducted a review of the literature related to drug impaired driving, and uncovered much new research that relates drug impairment to behavioral measures.

References

Adler, E.V., Burns, M., 1994. Drug recognition Expert (DRE) validation study. Final report to the Arizona Governor's Office of Highway Safety.

Bigelow, G.E., Bickel, W.E., Roache, J.D., et al., 1985. Identifying types of drug intoxication: laboratory evaluation of a subject examination procedure. U.S. Department of Transportation, National Highway Traffic Safety Administration, Report No. HS 806 753, Washington D.C.

Compton, R.P., 1986. Field evaluation of the Los Angeles Police Department evaluation drug detection procedure. U.S. Department of Transportation, National Highway Traffic Safety Administration, Report No. HS 807 012, Washington D.C.

Evans, S.M., Troisi, J.R., Griffiths, R.R., 1994. Tandospirone and alprazolam: comparison of behavioral effects and abuse liability in humans. J. Pharmacol. Exp. Ther. 271 (2), 683–694.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Heishman, S.J., Singleton, E.C., Crouch, D.J., 1996. Laboratory validation study of Drug Evaluation and Classification Program: ethanol, cocaine, and marijuana. J. Anal. Toxicol. 20, 466–483.

Heishman, S.J., Singleton, E.C., Crouch, D.J., 1998. Laboratory validation study of Drug Evaluation and Classification Program: alprazolam, D-amphetamine, codeine, and marijuana. J. Anal. Toxicol. 22, 503–514.

Jones, R.K., Shinar, D., Walsh, M.J., 2003. State of knowledge of drug impaired driving. U.S. DOT, National Highway Traffic Safety Administration, Report DOT HS 809 642, Washington D.C.

Kelly, T.H., Foltin, R.W., Serpick, E., Fischman, M.W., 1997. Behavioral effects of alprazolam in humans. Behav. Pharmacol. 8 (1), 47–57.

Kosnoski, E.M., Yolton, R.L., Citek, K., Hayes, C.E., Evans, R.B., 1998. The drug evaluation classification program: using ocular and other signs to detect drug intoxication. J. Am. Optom. Assoc. 69 (4), 211–227.

Linnoila, M., Hakkinen, S., 1974. Effects of diazepam and codeine, alone and in combination with alcohol, on simulated driving. Clin. Pharmacol. Ther. 15 (4), 368–373.

NHTSA Drug Evaluation and Classification Program, 1991. Briefing paper. National Highway Department of Transportation, U.S. Department of Transportation, Washington D.C.

Perez-Reyes, M., Hicks, R.E., Bumberry, J., Jeffcoat, A.R., Cook, C.E., 1988. Interaction between marihuana and ethanol: effects on psychomotor performance. Alcohol Clin. Exp. Res. 12 (2), 268–276.

Perez-Reyes, M., White, W.R., McDonald, S.A., Hicks, R.E., 1992. Interaction between ethanol and dextroamphetamine: effects on psychomotor performance. Alcohol Clin. Exp. Res. 16 (1), 75–81.

Preston, K.L., Wolf, B., Guarino, J.J., Griffiths, R.R., 1992. Subjective and behavioral effects of diphenhydramine, lorazepam, and methocarbamol: evaluation of abuse liability. J. Pharmacol. Exp. Ther. 262 (2), 707–720.

Schechtman, E., Shinar, D., 2005. Modeling drug detection and diagnosis with the 'drug evaluation and classification program'. Accid. Anal. Prevent. 37 (5), 852–861.

Shinar, D., Schechtman, E., 1998. Modeling the DRE evaluation of signs and symptoms to improve the validity of drug impairment diagnosis. Final Report submitted to the National Safety Council and the U.S. Department of Transportation, National Highway Traffic Safety Administration.

Smith, J.A., Hayes, C.E., Yolton, R.L., Rutledge, D.A., Citek, K., 2002. Drug recognition expert evaluations made using limited data. Forensic Sci. Int. 130, 167–173.

Stuster, J., Burns, M., 1998. Validation of the Standardized Field Sobriety Test battery at BACs below 0.10 percent. U.S. Department of Transportation, Report No. DOT HS 808 839, Washington D.C.

Sutton, L.R., 1983. The effects of alcohol, marihuana and their combination on driving ability. J. Stud. Alcohol 44 (3), 438–445.

Tedeschi, G., Bittencourt, P.R., Smith, A.T., Richens, A., 1983. Effect of amphetamine on saccadic and smooth pursuit eye movement. Psychopharmacology 79 (2–3), 190–192.

Townsend, T.N., Lane, J., Dewa, C.S., Brittingham, A.M., 1998. Driving after drug or alcohol use: Findings from the 1996 National Household Survey on Drug Abuse. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, DHHS Publication No. (SMA) 99 3273, Washington D.C.

Ward, N.J., Dye, L., 1999. Cannabis and driving: a literature review and commentary. DETR Road Safety Research 1999 Report No. 12.