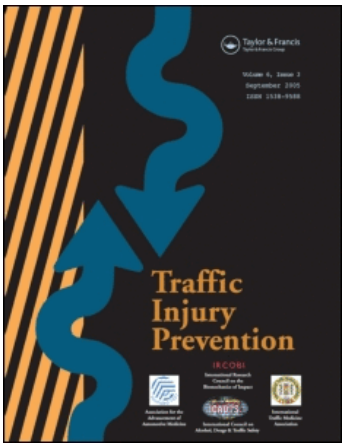

Traffic Injury Prevention
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713456148

Evaluation of the Drug Evaluation and Classification Program: A Critical Review of the Evidence

Douglas J. Beirness [a]; Jacques LeCavalier [a]; Deanna Singhal [b]
[a] Canadian Centre on Substance Abuse, Ottawa, Ontario, Canada [b] Traffic Injury Research Foundation, Ottawa, Ontario, Canada

Online Publication Date: 01 December 2007

To cite this Article Beirness, Douglas J., LeCavalier, Jacques and Singhal, Deanna(2007)'Evaluation of the Drug Evaluation and Classification Program: A Critical Review of the Evidence',Traffic Injury Prevention,8:4,368 — 376
To link to this Article: DOI: 10.1080/15389580701525651
URL: http://dx.doi.org/10.1080/15389580701525651

Traffic Injury Prevention, 8:368–376, 2007

ISSN: 1538-9588 print / 1538-957X online
DOI: 10.1080/15389580701525651

Taylor & Francis
Taylor & Francis Group

Evaluation of the Drug Evaluation and Classification Program: A Critical Review of the Evidence

DOUGLAS J. BEIRNESS and JACQUES LeCAVALIER
Canadian Centre on Substance Abuse, Ottawa, Ontario, Canada

DEANNA SINGHAL
Traffic Injury Research Foundation, Ottawa, Ontario, Canada

Objective. *A critical review of the existing evaluation studies on the Drug Evaluation and Classification (DEC) program was conducted to determine the validity and accuracy of the technique for identifying drivers under the influence of drugs.*

Methods. *Studies were divided into two categories—laboratory studies and field (i.e., enforcement) studies. A classification process was devised using common criteria based on the toxicology findings (i.e., drug positive or drug negative) and the opinion of the police officer who assessed the driver (i.e., drug positive or drug negative). A series of standard measures (Sensitivity, Specificity, False Alarm Rate, Miss Rate, Corroboration, and Accuracy) were calculated for each to assess the effectiveness of the DEC program.*

Results. *Laboratory studies do not provide overwhelming support for the accuracy with which officers trained in the DEC program can detect and identify the particular class(es) of drug involved based on psychophysical assessment alone. The detection and identification of the relatively low levels of drugs administered were typically better than chance but many cases were missed. The fact that some drugs were detected with greater accuracy than others suggests that the effects of these substances were more prominently manifested in the symptomology assessed by the DEC procedure. Although field enforcement studies are not as scientifically rigorous as laboratory studies, DEC assessments in an enforcement context have the benefit of information obtained from the arresting officer and from interviews with the suspect. In addition, the drug doses consumed by users are typically much higher than those permitted in controlled laboratory studies. In general, officers trained in the DEC program are able to identify persons under the influence of drugs and to specify the drug class responsible with a degree of accuracy that not only exceeds chance, but in some cases reaches a very high level.*

Conclusions. *There remains room for improvement in the DEC program. As further research becomes available, either from laboratory or field investigations or both, it needs to be incorporated into the program to enhance its validity and accuracy.*

Keywords Drugged Driving; Drug Evaluation and Classification (DEC); Drug Recognition Expert (DRE); Impaired Driving; Impairment Testing

In recent years, concern about drug use by drivers has become an increasingly more prominent social and road safety issue. Although there is a natural tendency to approach this issue with the same strategies and tactics that proved so successful with the alcohol-crash problem, it quickly becomes apparent that detecting and measuring drug use among drivers is considerably more complex than for alcohol. Most drugs cannot be detected in breath, and require toxicological analysis of samples of bodily fluids such as blood, urine, or oral fluid. This complication required the development of a different approach, one that would allow police officers to identify drivers under the influence of drugs in an efficient and effective manner.

During the 1970s, the Los Angeles Police Department utilized knowledge about the known signs and symptoms associated with the ingestion of various drugs to begin the development of a systematic procedure to evaluate drivers suspected of being under the influence of drugs. The procedure, which continues to evolve, has come to be known as the Drug Evaluation and Classification (DEC) program and is supported by the National Highway Traffic Safety Administration (NHTSA)

Received 6 December 2006; accepted 20 June 2007.

Address correspondence to D.J. Beirness, Canadian Centre on Substance Abuse, 75 Albert St., Suite 300, Ottawa, Ontario K1P 5E7. E-mail: dbeirness@ccsa.ca

Initial work on this project was conducted while the senior author was with the Traffic Injury Research Foundation (TIRF). The views expressed in this article are those of the authors and do not necessarily reflect those of their respective institutions or sponsors.

and coordinated by the Highway Safety Committee of the International Association of Chiefs of Police (IACP). In 1992, the committee adopted standards specifying the requirement for certification and re-certification of Drug Recognition Experts (DREs) and DRE instructors, standards for decertification and reinstatement of DREs, and standards for agency participation (IACP, 1999).

The DEC program is a systematic and standardized procedure, which involves a series of physical and psychomotor tests and concludes with the toxicological examination of a bodily fluid sample. The purpose of the procedure is to provide the officer with the necessary evidence to determine whether or not the suspect is impaired, whether the observed impairment is due to drugs, and which category (or categories) of drugs might be responsible. The results of the 12-step protocol, when corroborated by toxicological evidence of drug use, provide sufficient evidence to proceed with drug-impaired driving charges.

As the use of the DEC program expands across the United States, and into Canada, Europe, and Australasia, it is important to understand the scientific basis for the validity and accuracy of the technique. This article provides a critical review of the existing evaluation studies of the DEC program. The purpose is to provide a comprehensive and balanced assessment of the evidence of the validity of the program.

METHODS

Evaluation studies were divided into two categories—laboratory studies and field (i.e., enforcement) studies. These two types of studies differ not only in their methods but in their purpose. Laboratory studies are systematic investigations conducted with volunteers who consume measured doses of specific drugs and are then evaluated by police officers who have been certified as Drug Recognition Experts (DREs). Field studies typically involve an examination of data collected from suspected drug-impaired drivers within an enforcement setting. Both types of studies compare the judgements made by DREs concerning the use of drugs with objective evidence of drug use—either the type of drug administered (laboratory studies) and/or the results of toxicological tests for the presence of psychoactive substances (field studies). The purpose is to determine the extent to which the DEC procedure is effective in identifying the type of drug(s) responsible for the symptoms observed.

Measures of Accuracy

Most studies attempt to determine the degree of correspondence between the opinion of the DRE and the actual use of psychoactive substances. Several different measures are reported, each of which varies according to the criteria for "matches" or "hits" as well as the denominator used. To facilitate comparisons among studies, an effort was made to calculate a standard set of measures based on a comparison of the toxicology results (i.e., drug positive or drug negative) and the opinion of the DRE (i.e., drug positive or drug negative). Where possible, these measures were calculated for each drug class examined.

Table I Overall classification of DRE opinions against dose condition (data from Bigelow et al., 1985)

	DRE opinion		
Dose condition	Drug positive	Drug negative	
Drug positive	144 (TP)	125 (FN)	269
Drug negative	13* (FP)	38 (TN)	51*
	157	163	**320**

*Includes 11 cases where the drug administered was not the drug identified by the DRE. Sensitivity = 144/269 = .535; Specificity= 38/51 = .745; False alarm rate = 13/51 = .255; Miss rate = 125/269 = .465; Corroboration = 144/157 = .917; Accuracy = (144 + 38)/320 = .569.

The number of drug positive cases correctly identified by the DRE are considered "True Positives (TP)." Drug negative cases correctly identified by the DRE are labeled "True Negative (TN)." Mismatches between the DRE opinion and the toxicology are labeled either "False Positive (FP)" or "False Negative (FN)."

A simple classification table (such as that shown in Table I) provides the basis for six measures that can be used to assess the accuracy of the DEC program. The first measure derived from this table is ***sensitivity***—also known as the "hit" rate, or true positive rate. Sensitivity addresses the question "*Given that a driver has ingested a certain substance, how likely is it that the person will be detected by the DEC procedure?*" This measure is defined as the number of drug positive cases correctly identified by the DRE (**TP**) divided by the total number of drug positive cases identified by the toxicology (**TP+FN**). It is desirable to have a procedure that maximizes sensitivity. Tests with high sensitivity minimize false negatives.

The second important measure is the ***specificity*** or correct rejection rate. This measure is the proportion of all cases correctly identified by the DRE as not being under the influence of drugs. Specificity is the number of cases the DRE specifies as being drug negative (**TN**) divided by the total number of drug negative cases identified by the toxicology **(TN+FP)**. This measure is often not reported in field studies because there are typically few drug negative cases that are subjected to a DEC assessment. It is desirable to have a procedure with high specificity. Tests with high specificity minimize false positives.

The complements of sensitivity and specificity are the ***false alarm rate*** and ***miss rate***, respectively. The *false alarm rate* is the proportion of all drug negative cases **(FP+TN)** in which the DRE indicates the individual is under the influence of a drug **(FP)**. From a criminal justice perspective it is desirable to minimize the false alarm rate because it represents the likelihood an individual will be falsely accused by the DEC procedure.

The ***miss rate*** is the proportion of all drug positive cases that are judged by the DRE to be drug free. The miss rate is represented by **(FN/TP+FN)**. An optimal procedure should have a minimal miss rate.

An alternative measure that is often reported is the ***corroboration rate***, or ***positive detection rate***. This measure is defined

as the proportion of all persons identified by the DRE procedure as being under the influence of a given substance that are subsequently confirmed by the toxicology as being correctly identified. This measure answers the question "*How often are DREs correct in their judgements of the class of drug ingested by suspected impaired drivers?*" The corroboration rate is defined as **TP/(TP+FP)**.

The final measure, which will be referred to as ***accuracy***, is the proportion of all cases that are either correct "hits" or "correct rejections"—i.e., **(TP+TN)/(TP+FP+TN+FN)**. This measure reflects the overall proportion of cases correctly identified by the DEC procedure.

Wherever possible, these measures have been calculated for each of the studies reviewed. The value of these measures may differ from those presented in the published reports. In some cases, it was not possible to generate these measures, or certain assumptions were made about the available data to allow these measures to be generated. Sometimes, the measures were available on a drug-specific basis; other times, the only data available were for all drug classes combined.

RESULTS

Laboratory Studies

Experimental laboratory studies allow researchers to impose strict conditions on the type(s) of drugs ingested, the time elapsed between drug administration and behavioral assessment, and the variables collected from the assessment procedure. These studies examine the validity of the DEC procedure under ideal and controlled—albeit artificial—conditions. Using the same group of DREs and volunteers repeatedly over several sessions reduces the extent of variation attributable to individual differences among DREs, thereby providing a precise examination of the DEC procedure.

Laboratory studies can include a placebo condition, where volunteers are given an inactive substance as a means to examine the accuracy with which the DEC procedure is able to detect subjects who are *not* under the influence of any substance. Experimental studies often use a "double blind" procedure, in which neither the volunteer nor the DRE doing the evaluation are aware of what drug has been ingested, and restrict the questioning of volunteers to eliminate the bias introduced by admissions of drug use. These procedures provide a rigorous test of the psychophysical assessment, unbiased by prior knowledge of the type of drug ingested.

Bigelow et al. (1985). The original laboratory evaluation of the set of assessment procedures that came to be known as the DEC program was conducted at Johns Hopkins University School of Medicine (Bigelow et al., 1985). Eighty volunteers were randomly assigned to one of eight drug conditions (d-amphetamine 15 mg, d-amphetamine 30 mg, marijuana 12 puffs of 1.3% THC, marijuana 12 puffs of 2.8% THC, diazepam 15 mg, diazepam 30 mg, secobarbital 300 mg, or placebo) and were subsequently examined independently by each of four experienced DREs from the Los Angeles Police Department, yielding a total of 320 assessments (80 subjects by four raters). DREs were informed that some subjects would receive no active drug (i.e., a placebo condition) and that there would be no alcohol, PCP, LSD, or drug combinations. DREs were instructed to indicate a drug class even if they were not as confident about the judgment as they would normally be in an actual field situation. The time available for each evaluation was restricted to 20 minutes and the interview with the subject, the search for physical evidence, and breath alcohol test were eliminated from the procedure. Subjects were told not to volunteer any information about the type of drug they thought they might have received, eliminating cues associated with admissions or confessions of drug use.

Table I presents the measures of accuracy for all drugs combined. For all drug and dose conditions combined, the sensitivity of the DEC procedure was .535—i.e., the DEC procedure correctly identified 53.5% of all individuals who were given an active drug. The specificity was .745, indicating that the DEC procedure correctly identified 74.5% of individuals who were not under the influence of drugs. This figure, however, includes 11 cases where the drug administered was not the drug identified by the DRE. Of the 40 cases that were actually given a placebo, DREs correctly identified 38 (95%) as not being under the influence of a drug.

The false alarm rate indicates that 25.5% of cases were identified as being under the influence of a drug when in fact they were not. Again, this figure includes 11 cases where the DRE identified the wrong drug. Exclusion of these 11 cases results in a false alarm rate of 4%. The miss rate (46.5%) indicates that almost half of all drug positive cases were missed by the DEC procedure. To a large extent, this can be attributed to the low-dose condition in which volunteers were administered a dose that would not necessarily be expected to have profound and observable effects, particularly on persons with a history of drug use. In cases where the DRE determined that the individual was under the influence of a drug, the opinion was correct (i.e., corroboration rate) 92% of the time.

From the data provided in the report, it is also possible to calculate the various measures for each drug and drug category. These measures are presented in Table II. The same placebo cases have been used repeatedly in the calculations for each drug category. Also, the actual number of cases for a specific drug or drug class is relatively small, reducing the overall reliability of the figures.

Table II Measures of DEC accuracy by drug and drug class (data from Bigelow et al., 1985)

	Measure					
Drug/class	Sensitivity (%)	Specificity (%)	False alarm (%)	Misses (%)	Corroboration (10%)	Accuracy (10%)
Stimulants	20.0	86.4	13.6	80.0	72.7	43.5
Cannabis	48.8	92.7	7.3	51.2	92.9	63.6
Depressants	74.2	84.4	15.6	25.8	92.7	77.0
Diazepam	63.8	84.4	15.6	36.2	87.9	71.2
Secobarbital	95.0	84.4	15.6	5.0	84.4	92.9

It is apparent that the DREs are better at detecting some drugs than others. For example, the sensitivity rate for CNS stimulants indicates that only 20% of subjects who were administered a stimulant were correctly identified. Stimulants also had the lowest corroboration rate of all drugs studied.

Two different CNS depressant drugs were administered in this study—diazepam and secobarbital. Table II shows the results for these drugs separately as well as combined as a drug class. Taken together, DREs had the greatest accuracy in identifying individuals under the influence of these types of drugs—almost three quarters of subjects who were administered one of these drugs were correctly identified by DREs as being under the influence of this class of drugs. The accuracy rate was highest for those administered secobarbital.

Heishman et al. (1996). More than 10 years after the first laboratory investigation of the DEC program, two other laboratory studies were conducted. The first study examined the administration of set doses of ethanol, cocaine, and marijuana (Heishman et al., 1996); the second administered doses of alprazolam (a CNS depressant), d-amphetamine (a CNS stimulant), codeine (a narcotic analgesic), and marijuana (Heishman et al., 1998). Both studies also included a placebo condition in which no active drug was administered. The primary goal of these studies was to determine the validity of the measures collected as part of the DEC procedure in predicting specific categories of drug use; a secondary goal was to determine the accuracy of DREs in predicting the category of drug used.

In the first study, 18 experienced drug user volunteers from the community were recruited to participate in nine experimental sessions during which they received placebo, a low dose, or a high dose of either ethanol, cocaine (a CNS stimulant), or cannabis for a total of 162 experimental sessions. In four of these sessions, the drug (cannabis) could not be detected in confirmatory blood samples and were eliminated from the analyses. A double blind drug administration procedure was followed.

Twenty-eight certified DREs from eight states were used over the course of the study to evaluate dosed participants. DREs were not permitted to interview subjects but were informed that participants may have received ethanol and/or one of five classes of drugs (CNS depressants, CNS stimulants, phencyclidine, narcotic analgesics, or cannabis) or no active drug at all. In fact, with the exception of placebo sessions, only one active drug was ingested in each session.

Table III presents the six measures of accuracy for each of the three drugs examined as well as for all three drug classes combined. The measures of accuracy in Table III do not necessarily agree with those presented in the original report. The discrepancy is a consequence of our classifying as errors those cases in which the participant was given an active drug but judged to be "not impaired" by the DRE. This approach is more conservative than that employed by the authors who included in their analyses only those cases in which the DRE determined the subject to be impaired.

Cocaine proved to be the most difficult substance to detect consistently. Only about half of all cannabis cases were accurately identified. The figures presented for alcohol should be viewed with caution given the relatively low doses of alcohol administered.

Table III Measures of DEC accuracy by drug and drug class (data from Heishman et al., 1996)

	Measure					
Drug	Sensitivity (%)	Specificity (%)	False alarm (%)	Misses (%)	Corroboration (%)	Accuracy (%)
Cocaine	13.2	61.1	38.9	86.1	41.4	29.6
Cannabis	53.1	61.1	38.9	46.9	70.8	56.0
Alcohol	41.7	55.6	44.4	58.3	65.2	46.3
Combined	35.6	59.3	40.7	64.4	62.7	43.7

Heishman et al. (1998). The second study by Heishman and colleagues (1998) examined the accuracy of the DEC procedure in identifying volunteers who had been administered standard doses of either alprazolam (a CNS depressant), d-amphetamine (a CNS stimulant), codeine (a narcotic analgesic), or marijuana. The 4 drugs were examined in concurrent experiments, each involving 12 volunteers. Each volunteer participated in 6 experimental sessions. On each session, they ingested either a zero dose (i.e., placebo), a low dose, or a high dose. Participants were evaluated by one of 28 trained and certified DREs. The DEC procedures were modified to exclude questions about recent drug use or admissions about drug use.

Overall, the authors report that 76% of DRE decisions were consistent with drug use but in only 32% of cases was the DRE judgement of drug class consistent with the actual drug administered. From the re-analysis of the data for specific drugs presented in Table IV, it is evident that none of the substances was detected reliably (i.e., low sensitivity) by the DEC procedure. However, the doses of the various drugs administered were relatively modest and, with the exception of cannabis, were selected so as not to exceed the maximum daily therapeutic dose. The effects of these low-dose levels would not be expected to be large. This was reflected in the finding that approximately half of all cases were deemed "not impaired." In many of these "not-impaired" cases, however, the officer noted signs and symptoms of impairment and indicated the drug or drug categories believed responsible for these symptoms. The drug categories listed by the DRE for these "unimpaired" cases were not reported in the original paper but have been considered as "misses" in our

Table IV Measures of DEC accuracy by drug and drug class (data from Heishman et al., 1998)

	Measure					
Drug	Sensitivity (%)	Specificity (%)	False alarm (%)	Misses (%)	Corroboration (%)	Accuracy (%)
Alprazolam	27.1	62.5	37.5	72.9	59.1	38.9
Cannabis	30.4	59.1	40.9	69.6	60.9	39.7
Amphetamine	4.2	79.2	20.8	95.8	28.6	29.2
Codeine	14.9	82.6	17.4	85.1	63.6	37.1

Table V Measures of DEC accuracy by drug and drug class (data from Shinar & Schechtman, 2005)

	Measure					
Drug	Sensitivity (%)	Specificity (%)	False alarm (%)	Misses (%)	Corrob[1] (%)	Accuracy[2] (%)
Alprazolam	47	80	20	53	*30.6*	*39.1*
Cannabis	49	69	31	51	*42.9*	*41.7*
Amphetamine	10	91	9	90	*36.7*	*41.1*
Codeine	45	72	28	55	*7.8*	*31.4*
All drugs	72	43	57	28	71	62

[1]The numbers for individual drugs are based on matches with the first drug indicated by the DRE. These numbers are reported by Shinar and Schechtman (2005) in their Table II (p. 847).

[2]Accuracy for each drug has been calculated using all 44 matches for the 102 placebo cases.

reanalysis of these data. Hence, the figures in Table IV are most likely underestimates of the accuracy of the DEC procedure.

Even with the limitations noted previously, it remains apparent in Table IV that the accuracy of detection varied according to the drug administered. Participants who were dosed with cannabis or a CNS depressant (alprazolam) were more likely to be identified accurately than those under the influence of either amphetamine or codeine.

Shinar and Schechtman (2005). These authors re-examined the data from Heishman et al. (1998) and included the DRE opinions concerning the suspected drug class in those cases that were deemed "not impaired."

Table V presents the six measures used to assess the accuracy of DRE judgments concerning the drug consumed based on the officer's assessment. For individual drugs, the last two measures—corroboration and accuracy—were not reported in the paper and could only be calculated on the basis of the first drug category listed by the DRE. Thus, if the DRE listed both a CNS depressant and cannabis as the drug classes, the case was only considered a match if the actual drug ingested was a CNS depressant. The other measures were calculated by considering a case to be a match if the drug administered was noted as either the first or second drug category listed by the DRE. Thus, the measures of corroboration and accuracy calculated from the data in the paper must be considered conservative estimates.

Overall, the officers' records of making correct decisions concerning the detection of impairment given that the subject had ingested a particular drug was modest—i.e., sensitivity of 72%. Less than half of cases not administered any active drug were identified as being unimpaired. The relatively high false alarm rate (57%) may have been, in part, a result of the officers' beliefs that most participants had ingested drugs.

Field Studies

Field studies of the DEC program involve an examination of data collected in an enforcement setting. These studies compare DRE judgements of suspected drug use by persons arrested for an impaired driving offence with the results of toxicological tests for the presence of psychoactive substances. By definition, field studies lack the controls imposed in laboratory studies. The lack of controls can both help and hinder the observed accuracy with which DREs identify impairment and the drug class(es) responsible. For example, whereas laboratory studies use volunteers who have been carefully screened and administered known quantities of one specific substance (or placebo), field studies involve real drivers who, as a result of some event—collision, spot check, erratic driving—have come to the attention of a police officer who has reason to believe they might be under the influence of drugs. Although the range of possible drugs used by real drivers is vast, the DRE has every reason to suspect that some substance has been used and, based on experience, may have reason to believe that certain classes of drugs may be more common than others.

The quantities of drug(s) ingested by drivers suspected of impaired driving can be considerably larger than those used in laboratory studies. Higher doses create more profound effects that are easier to detect. However, drugs are often used in combination with other drugs and/or alcohol, a condition not examined in laboratory studies, which can mask some symptoms and enhance others, making it more difficult to identify the particular substances involved.

On the other hand, whereas laboratory studies often do not allow the DRE to conduct an interview with the subject and ask about substance use, this is a key feature of the DEC protocol used in field settings that can provide officers with valuable information and may lead to a confession about drug use.

In sum, there are many issues that distinguish the two types of studies that can lead to differences in the findings. This section examines the findings from relevant field studies.

Compton (1986). The first field evaluation of the DEC program was conducted by the National Highway Traffic Safety Administration (NHTSA) in Los Angeles (Compton, 1986). The study examined adults suspected of driving under the influence of drugs or a combination of alcohol and drugs during the summer of 1985. Drivers involved in collisions were excluded. Of the initial 219 suspects identified, 18 were determined to be "not impaired" and were excluded from the study sample. Of those remaining, 173 (86%) agreed to provide a blood sample. Suspects performed a standardized field sobriety test (SFST) at roadside and were brought to one of two jail facilities for drug evaluation by one of 25 DREs selected for this project. Breath alcohol tests were performed prior to the DEC assessment and a blood sample was taken within two hours of arrest.

The most commonly detected drugs were PCP (56%), alcohol (53%), and cannabis (45%). More than one drug was found in 72% of suspects; half of all suspects were also positive for alcohol. In ten cases, alcohol was the only substance found; in one case, no drugs or alcohol were present. At least one of the drugs identified by the DREs as being responsible for the impairment was confirmed by toxicology in 87% of cases. This figure excludes alcohol as one of drugs because breath test results were known to DREs. Opinions about the specific drug(s) involved

Table VI Measures of DEC accuracy by drug class (data from Compton, 1986)

	Measure					
Drug	Sensitivity (%)	Specificity (%)	False alarm (%)	Misses (%)	Corrob (%)	Accuracy (%)
PCP	90.7	89.5	10.5	9.3	91.7	90.2
Cannabis	59.7	86.4	13.5	40.3	78.0	74.6
Opiates	65.4	97.9	2.1	34.6	85.0	93.1
Stimulants	19.0	94.7	5.3	81.0	33.3	85.5
Depressants	73.7	90.9	9.1	26.3	50.0	89.0
All classes	70.4	92.6	7.4	29.6	78.6	86.5

Table VII Measures of DEC accuracy by drug class (data from Preusser et al., 1992)

	Measure					
Drug	Sensitivity (%)	Specificity (%)	False alarm (%)	Misses (%)	Corrob (%)	Accuracy (%)
PCP	75.3	98.4	1.6	24.7	70.5	97.2
Cannabis	78.4	73.2	26.7	21.6	68.4	75.4
Opiates	75.9	94.3	5.7	24.1	67.3	91.9
Stimulants	57.4	84.9	15.1	42.6	68.0	75.1
Depressants	68.6	86.4	13.6	31.4	48.2	83.6

were more likely to be entirely correct (all identified drugs found in blood) in cases where only one or two drugs were detected in blood. As the number of drugs found in suspects' blood increased, the greater the likelihood the DRE would get at least one of them correct.

Table VI presents the six measures of DRE accuracy in detecting drugs in suspected impaired drivers. All 173 cases were used to generate the measures for each drug class. This approach tends to overestimate specificity (and underestimate false alarms) because all cases that test negative for a particular drug class are included even though they may be positive for some other drug class. The measures for "All classes" are calculated on the basis of all possible drug detections for each case—i.e., 173 cases × 5 drug classes = 865. Hence each case has been included five times.

As was the case in the laboratory studies, DRE accuracy varied according to drug class. PCP was not only the most commonly found drug but it was correctly identified by the DRE in 91% of cases where the blood test was positive for PCP. In only 10.5% of cases did the DRE indicate that PCP was present and it was not confirmed by the toxicology (false alarms). Cannabis was frequently found but DREs had somewhat more difficulty identifying it—only 59.7% of cases that tested positive for cannabis were correctly identified by the DREs. Stimulants proved to be the most difficult to detect. Of the 21 cases in which stimulants (exclusively cocaine) were detected in blood, DREs correctly identified only four (19%).

Overall, when DREs identified a drug as being present, their opinion was confirmed by blood tests 79% of the time (corroboration rate). Unconfirmed suspicions of drug use (i.e., false alarms) were minimal (7.4%). DREs missed approximately 30% of all drug positive cases.

Preusser et al. (1992). In one of the largest field studies to date, Preusser et al. (1992) examined the records of 1,842 cases evaluated by DREs in five states and compared the drug class determined by the DRE with the toxicology results. Of the 1,842 cases available for analysis, 1,711 (92.9%) were deemed by the DRE to be under the influence of one or more drugs. The remaining cases were excluded because of medical reasons, impairment was not evident, or the suspect refused to participate.

From this pool of 1,711 cases, 1,469 had toxicological test results available. At least one drug was found in 1,236 (84.1%) cases. Overall lab tests confirmed the presence of the drug named by the DRE in 64.1% of cases.

Table VII displays the measures of DRE accuracy for the most commonly found drug categories. Hallucinogens and inhalants have been excluded due to the extremely low number of cases. The figures presented in the table are based on all 1,469 cases. For each drug category, cases in which the DRE did not specify the particular drug and the toxicological test confirmed it was not present were considered "True Negatives." This tends to inflate the specificity and underestimate the false alarm rate for individual drugs.

Once again, these figures reveal differences in the ability of DREs to detect specific drug categories. These differences are somewhat smaller than in other studies, most likely as a consequence of the larger number of cases available for analysis. Nevertheless, stimulants appear to be the most difficult to detect accurately and cannabis is responsible for the greatest proportion of false alarms.

Hardin et al. (1993). A field evaluation of the DRE program in Minnesota compared the judgements of officers with urine samples collected from 76 cases (Hardin et al., 1993). Cannabis was the most commonly found substance, accounting for 68% of positive samples. Narcotics (14%), stimulants (9%), and depressants (9%) made up the rest. There were five cases in which the DREs determined the suspect to be "not under the influence." These latter cases were excluded from further consideration by the authors.

Table VIII displays the six measures of DRE accuracy calculated from the tables provided by the authors. The figures for each individual drug class are based on all 71 cases in which the DRE determined the suspect to be under the influence. The figures for all drug classes combined are based on the original 76 cases, including those that were deemed "not under the influence."

Overall, the DREs' judgements were confirmed by the toxicology in 87% of cases. DREs were able to detect drugs when present in 92% of cases. The proportion of drug-positive cases that were missed was quite low (7.7%). The ability to detect drug negative cases appeared to be poor (9.1%) but there were only 6 cases, 3 of which had positive blood alcohol levels.

Even within this relatively small sample of cases, the accuracy with which individual drug classes were detected varied

Table VIII Measures of DEC accuracy by drug class (data from Hardin et al., 1993)

	Measure					
Drug	Sensitivity (%)	Specificity (%)	False alarm (%)	Misses (%)	Corrob (%)	Accuracy (%)
Cannabis	93.8	82.6	17.4	6.2	91.8	90.1
Depressants	69.2	91.4	8.6	30.8	64.3	87.3
Stimulants	37.5	94.5	5.4	62.5	66.7	81.6
Opiates	66.7	90.3	9.7	33.3	50.0	87.3
All drugs	92.3	9.1	90.9	7.7	86.9	80.3

Table X Measures of DEC accuracy by drug class (data from Smith et al., 2002)

	Measure					
Drug	Sensitivity (%)	Specificity (%)	False alarm (%)	Misses (%)	Corrob (%)	Accuracy (%)
Cannabis	80.5	76.6	23.3	19.5	94.2	79.9
Stimulants	77.8	84.3	15.7	84.3	96.0	78.9
Depressants	68.9	93.7	6.3	31.1	97.7	73.9
Narcotics	94.0	83.1	16.9	6.0	94.4	91.3
All drugs	79.7	65.5	34.4	20.2	96.8	78.7

considerably. Cannabis was detected most accurately and the identification of stimulants was the least accurate.

Adler and Burns (1994). This study examined 500 DRE case records from Arizona over a 53-month period from 1989 through 1993. Of the 500 cases, bodily fluid samples were obtained from 484. Toxicological analysis of samples found one drug in 163 (33.7%) cases, two or more drugs in 253 (52.3%) cases, and no drugs in 68 (14.0%). Of those cases in which drugs were found, DREs correctly identified at least one drug in 378 (90.9%).

Cannabis was the most commonly detected substance, followed by cocaine and benzodiazepines. Approximately one third of all suspects had also consumed alcohol. Most of the BACs were relatively low; only 5% of all alcohol-positive cases had a BAC of .10% or higher.

Table IX shows the overall distribution of cases according to the opinion of the DRE concerning drug use and the toxicology results, along with the measures of accuracy. There was insufficient detail in the report to provide all these measures for each individual drug class.

Overall, the accuracy with which DREs were able to identify drugs that were indeed present was quite high (sensitivity = 90.9%). Similarly, the drug class identified by the DRE was subsequently confirmed by toxicology in 90% of cases. However, the false alarm rate (61.7%) was relatively high, indicating that DREs have a tendency to indicate drug involvement when no drugs are present.

The authors also report that many suspects admitted to drug use. Such admissions often arise when the DRE confronts the suspect with the findings from the evaluation. Although there can be questions about the veracity of admissions by suspects, admissions of drug use were verified by toxicology approximately 90% of time. Hence, suspects' statements regarding the types of drugs ingested are not only common, but appear to be confirm the opinion of the DRE most of the time.

Table IX Overall classification of DRE opinions against toxicology results (data from Adler & Burns, 1994)

	DRE opinion		
Toxicology	Drug positive	Drug negative	
Drug positive	378	38	416
Drug negative	42	26	68
	420	64	Total = 484

Sensitivity = 90.9%; Specificity = 38.2%; False alarm rate = 61.7%; Miss rate = 9.1%; Corroboration = 90.0%; Accuracy = 83.5%.

Smith et al. (2002). A re-analysis of a selected set of DRE evaluation reports from the state of Oregon was presented by Smith et al. (2002). In the 70 cases selected, the opinion of the DRE matched the drug(s) identified by toxicology, there was a complete DEC assessment, cases were representative of the various drug classes, and all had a zero BAC. Twenty cases involved cannabis, 19 stimulants, 14 depressants, and 12 narcotics. In 5 cases, no drugs were found. The field reports of these 70 cases were modified to remove specific portions—i.e., the toxicology results, breath test results, confessions, arresting officer statements—and were then sent to 18 DREs for re-evaluation. The purpose was to compare DRE assessments made by these 18 DREs using only the psychophysical data—i.e., the signs and symptoms of drug use—with the original assessments made using the full spectrum of information typically available to the DRE.

The re-evaluation of these 70 cases by each of 18 DREs yielded a total of 1,260 judgments for comparison with the toxicological data. Table X presents the six measures of accuracy for each of the five categories of drugs examined plus all drug categories combined. The 90 judgments of the five drug-negative cases were used to compute these measures for each drug category. Of the 90 drug-negative judgments, 59 correctly identified no drug impairment (specificity = 65.5%). The remaining 31 judgments were considered either misses or false alarms depending on the drug category being examined and the drug category identified by the DRE.

The results indicate that DREs are able to make judgements about drug use with a reasonable degree of accuracy using only the psychophysical symptoms reported by a third party and without benefit of a face-to-face interaction with suspects, breath alcohol data, or the results of interviews with the suspects. Although the cases selected might appear to represent "ideal" cases—i.e., no alcohol, single drug, original DRE opinion matched toxicology—the results nevertheless provide an indication of the interrater consistency (i.e., reliability) of DRE judgments.

As was evident in other studies, the data confirm that some substances are easier to detect than others. This would appear to suggest that the physical symptoms associated with some drugs are reasonably clear indicators of its presence. At the same time,

however, it is evident that errors can still be made. The authors suggested that these errors reflect the extent to which interviews with, and confessions of, suspects are important components of the DEC assessment process.

Other studies. There are two other field evaluations of the DRE program—one from Arizona (Adler, 1990), the other from Texas (Louie, 1990). The reports of these studies are brief and contain insufficient details to calculate the six measures of DRE accuracy that have been presented for the other studies. Both present a measure of accuracy that appears to be the corroboration rate—i.e., the percent of cases in which the judgment of drug class by the DRE was confirmed by the toxicology.

Adler (1990) presented data from 185 cases and reports an overall corroboration rate of 82.7%. For individual drug classes, the DRE judgments confirmed by toxicology were PCP 93.8%, cannabis 87.5%, stimulants 75.5%, narcotics 72.7%, and depressants 62.1%.

Louie (1990) presented data on 100 cases from the Texas DEC program. Overall, DREs correctly identified the correct drug category (as determined by toxicology) in 70.7% of cases. For individual drug classes, the corroboration rates were PCP 60%, opiates 100%, cannabis 63.6%, depressants 50%, and stimulants 82.6%.

DISCUSSION

The evidence from experimental and field evaluations of the DEC program provide somewhat different perspectives on its validity and accuracy. Field studies provide stronger support for the program than experimental laboratory evaluations. Reconciling the different results requires an examination of a variety of factors that can influence the outcome of the two types of studies.

Laboratory studies are methodologically stronger than field studies because of the controlled conditions under which volunteer participants are tested.

At the same time, however, the experimental controls employed in a laboratory situation to enhance methodological rigour also create an artificial environment that differs considerably from field—i.e., enforcement—settings in which the DEC procedures are implemented. Several factors emerge as critical in this respect. First, DRE assessments in an enforcement context have the benefit of information obtained from the initial investigating officer and from interviews with the suspect. The fact that an individual has been presented to a DRE for assessment indicates that the individual has been detained for suspicion of driving under the influence of drugs and/or alcohol. In many cases, the suspect will have failed a field sobriety test and there may be physical evidence of drug use present. Oftentimes, the suspect will confess to the DRE during the interview or upon presentation of the evidence obtained from the assessment. Such statements constitute valuable evidence that often confirms the opinion of the DRE about the type of drug ingested.

A factor that complicates the DRE assessment in field settings is the common practice of poly-drug use. Whereas laboratory studies have restricted the dosing conditions to a single drug at a time without the concomitant use of alcohol, this would appear to be the exception rather than the rule in field settings. Toxicology tests often reveal that persons suspected of driving under the influence of drugs have ingested more than one substance. Many suspects also test positive for alcohol. Depending on the substances ingested, some symptoms may be masked; others may be inconsistent with any particular drug class. This complicates the task of the DRE and can lead to errors.

To some extent, the differences in results between laboratory and field studies can be attributed to the dose of drugs used. In laboratory settings, drug doses are strictly controlled and are typically within the "normal" or therapeutic range of the drug. Experienced drug users may have developed a tolerance to some of the effects of the drugs at the dosages administered, making it difficult to detect impairment. Persons arrested for suspicion of drug-impaired driving may also have developed a degree of tolerance to various substances. However, drug users most likely administer higher drug doses that may be more readily detectable by DREs.

There are many other methodological issues that can be identified to highlight the limitations of research—both laboratory and field studies—in this area. For example, the experience of the DREs used to evaluate cases, the type of fluid sample collected for toxicological analysis, the detection thresholds of substances in samples, the time the sample was drawn, and the procedures used to select or exclude cases are several of the issues that can influence the accuracy of the DEC procedure.

Despite all the known limitations of the studies examined, it is apparent that officers trained in the DEC program are able to identify persons under the influence of the drugs and to specify the drug class responsible with a degree of accuracy that not only exceeds chance, but in some cases reaches a very high level. From an enforcement perspective, it is important that the judgments of DREs concerning drug use are corroborated by toxicology in a high proportion of cases and that the proportion of false positive cases be minimal. For the most part, the available studies demonstrate this to be the case.

However, there remains a substantial proportion of drug-positive cases that are missed—or misspecified. These may be cases that did not manifest gross symptoms of drug use or the level of the drug ingested had waned in the interval between identification as a suspected impaired driver on the road and assessment by a DRE. In a field setting, DREs would not expect to see many drug-negative cases. Nevertheless, it is important for the overall validity and accuracy of the DEC program that when these cases are presented they are identified with a high degree of accuracy.

The evaluation studies clearly illustrate that the DEC program is not perfect. As detailed and objective as the procedures have been designed to be, it requires experience to conduct the assessment with the degree of precision required. There is also a subjective component involved in the process whereby the officer must make judgments about the extent to which an observation exceeds the standard of what might be considered "average" or

"normal." Hence, there are numerous opportunities for errors to creep into the process.

It is also apparent the accuracy of the DEC procedure varies according to drug class. In essence, some drug types are more difficult to identify than others. For example, it would appear the physical symptoms associated with the use of PCP are rather distinctive and are readily apparent in most cases. On the other hand, the use of stimulants is often missed by the DEC assessment or confused with other substances. The rates of detection of cannabis are reasonably high but suffer in cases where cannabis is used in combination of other substances, including alcohol. The findings suggest that improvements are needed in the procedures used to detect some drug classes.

In evaluating the accuracy of DRE assessments, it would be beneficial to know the level of the substance(s) found in the fluid sample rather than simply an indication of its presence or absence. Very low drug level(s) might help to explain some of the cases that are missed. The drug levels could also be used to identify thresholds for the detection of the various drugs by the DEC procedure.

A few studies have examined the individual data elements collected in the course of a DRE assessment in an attempt to identify the best predictors of impairment by the various types of drugs (Heishman et al., 1998; Schechtman & Shinar, 2005; Shinar & Schechtman, 2005). The results of these types of analyses show that some of the variables collected during an assessment are better indicators than others for identifying certain classes of drugs. Efforts to focus on these indicators could improve the accuracy of DRE judgments.

In addition, these studies suggest that the number of pieces of information collected during a DRE assessment (approximately 100) may well be too large to reasonably consider in rendering a decision about the class of drug involved. In fact, Shinar and Schechtman (2005) concluded that DREs apparently used only one or two "pivotal" signs/symptoms to guide their decision concerning drug category while ignoring others, even if contradictory to their judgment. Although this approach may simplify the task for DREs, it can lead to errors. The DEC program could be improved by providing officers with a more efficient means of analyzing and weighting combinations of signs and symptoms associated with various drugs. Whether this is best accomplished through more extensive training or the development of a computer algorithm that could be applied to the data collected remains to be determined.

Identifying drug-impaired drivers and determining the type of drug-responsible for the impairment is neither simple nor straightforward. The DEC program provides the police with a valid investigative tool to facilitate the enforcement of drug-impaired driving laws. Nevertheless, despite the generally positive results of the evaluation studies reviewed in this paper, there remains room for improvement in the DEC program. As further research becomes available, either from laboratory or field investigations or both, the knowledge needs to be incorporated into the program to further enhance its validity and accuracy.

ACKNOWLEDGMENTS

This work was funded by the Royal Canadian Mounted Police.

REFERENCES

Adler EV. (1990) Arizona's Drug Recognition Program: A Performance Assessment, *J. Southwestern Ass. For. Sci.*, Vol. 12(1), pp. 12–16.

Adler EV, Burns M. (1994) Drug Recognition Expert (DRE) Validation Study, Arizona Governor's Office of Highway Safety, Phoenix.

Bigelow GE, Bickel WE, Roache JD, Liebson IA, Nowowieski P. (1985) Identifying Types of Drug Intxoication: Laboratory Evaluation of the Subject Examination Procedure, DOT HS 806 753, National Highway Traffic Safety Administration, Washington, DC.

Compton RP. (1986) Field Evaluation of the Los Angeles Police Department Drug Detection Program, National Highway Traffic Safety Administration, Washington, DC.

Hardin G, Meyer RR, Jejuridar SC. (1993) Corroboration Study: A Comparison of DRE Opinions to Toxicology Evaluations of Impaired Drivers in Minnesota. Paper presented at the Joint Meeting of the Society of Forensic Toxicologists and the California Association of Toxicologists.

Heishman SJ, Singleton EG, Crouch DJ. (1996) Laboratory Validation Study of Drug Evaluation and Classification Program: Ethanol, Cocaine, and Marijuana, *J. Analyt. Tox.*, Vol. 20, pp. 468–483.

Heishman SJ, Singleton EG, Crouch DJ. (1998) Laboratory Validation Study of Drug Evaluation and Classification Program: Alprazolam, d-Amphetamine, Codeine, and Marijuana, *J. Analyt. Tox.*, Vol. 22, pp. 503–514.

International Association of Chiefs of Police. (1999) The International Standards of the Drug Evaluation and Classification Program, The DEC Standards Revision Subcommittee of the Technical Advisory Panel of the IACP Highway Safety Committee. IACP, Arlington, VA.

Jones RK, Shinar D, Walsh JM. (2003) State of Knowledge of Drug-Impaired Driving, DOT HS 809 642, National Highway Traffic Safety Administration, Washington, DC.

Louie P. (1990) Report on the Drug Recognition Evaluation (DRE) Program in Harris County, Texas, *J. Southwestern Ass. For. Sci.*, Vol. 12(1), pp. 17–18.

Preusser DF, Ulmer RG, Preusser, CW. (1992) Evaluation of the Impact of the Drug Evaluation and Classification Program on Enforcement and Adjudication, DOT HS 808 058, National Highway Traffic Safety Administration, Washington, DC.

Schechtman E, Shinar D. (2005) Modeling Drug Detection and Diagnosis with the Drug Evaluation and Classification Program, *Acc. Anal. Prev.*, Vol. 37, pp. 852–861.

Shinar D, Schechtman E. (2005) Drug Identification Performance on the Basis of Observable Signs and Symptoms, *Acc. Anal. Prev.*, Vol. 37, pp. 843–851.

Smith JA, Hayes CE, Yolton RL, Rutledge DA, Citek K. (2002) Drug Recognition Expert Evaluations Made Using Limited Data, *For. Sci. Int.*, Vol. 130(2/3), pp. 167–173.