



ELSEVIER

Drug and Alcohol Dependence 70 (2003) S41–S54

**DRUG and
ALCOHOL
DEPENDENCE**

www.elsevier.com/locate/drugalcdep

Principles of initial experimental drug abuse liability assessment in humans[☆]

Roland R. Griffiths^{a,b,*}, George E. Bigelow^a, Nancy A. Ator^a

^a Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, 5510 Nathan Shock Drive, Baltimore, MD 21224, USA

^b Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Abstract

This paper describes the rationale and procedures for conducting what is considered by many to be the current “gold standard” for initial abuse liability testing of a novel compound: the classic acute dose–effect comparison study in volunteers with histories of drug abuse. Such a trial is most appropriate for predicting the likelihood of abuse by drug abusers and, in turn, the extent of drug diversion and illicit street sales if the novel compound became available in the community. The dose–effect abuse liability trial typically involves a double-blind complete crossover design in 10–14 subjects with histories of polydrug abuse in a controlled clinical pharmacology laboratory setting. Drug conditions usually involve placebo, three doses of the novel compound and three doses of an appropriate reference compound of known abuse liability. In each session, the time-course of effects of a single drug dose are evaluated. Intervals between experimental sessions are typically 1 to several days. The importance of testing high supra-therapeutic doses of the novel drug for the validity of the trial is emphasized, and the use of a dose run-up pilot study for selecting maximal doses and matching doses between the novel and comparison compound is explained. The rationale and description of outcome measures is discussed, including measures that reflect likelihood of abuse (e.g. drug vs. money choice and subject ratings of liking, good effects, estimated monetary street value), secondary measures that should be considered in interpreting likelihood of abuse (e.g. drug identification, subject-rated side effects and mood changes), and additional concurrent measures to establish equivalence of the novel and comparison compound (e.g. behavioral performance, observer-rated assessments, physiological measures).

© 2003 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Drug abuse liability; Humans; Drug abusers; Drug development; Anxiolytics, hypnotics, drug liking; Multiple-choice procedure, ARCI, POMS

1. Introduction

The purpose of this paper is to describe the rationale and procedures for conducting what is considered by many to be the current “gold standard” for initial abuse liability testing of a novel compound: the classic acute dose–effect comparison study in volunteers with histories of drug abuse. The historical development of these methods for abuse liability testing has been discussed previously (Jasinski, 1977; Jaffe and Jaffe, 1989), and

there are two excellent monographs that provide a detailed presentation of various strategies for drug abuse liability assessment in humans (Fischman and Mello, 1989; Camí et al., 1991b). Also, focused reviews have addressed abuse liability assessment of specific drug classes: stimulants and anorectics (Foltin and Fischman 1991a,b), anxiolytics and hypnotics (Roache and Griffiths, 1989; Evans et al., 1991; de Wit and Griffiths, 1991), opioids and analgesics (Jasinski, 1977; Bigelow, 1991), NMDA antagonists and neuroprotectants (Klein et al., 1999).

Table 1 presents the primary features of an acute dose–effect comparison abuse liability trial. Typically, such a trial compares the profile of acute effects of a range of doses of the novel compound to those of placebo and a range of doses of a reference compound of known positive abuse liability on an array of outcome

[☆] Based on a background paper prepared for the October 28–29, 2002 College on Problems of Drug Dependence sponsored meeting on Assessing the Abuse Liability of CNS Drugs held in Bethesda, MD.

* Corresponding author. Tel.: +1-410-550-0034; fax: +1-410-550-0030.

E-mail address: rgriff@jhmi.edu (R.R. Griffiths).

Table 1
Features of the classic acute dose–effect abuse liability trial

-
- 1) Typical trial design characteristics
 - A complete crossover design in 10–14 subjects
 - Single doses of drug are evaluated over a period of several minutes to several hours, depending on the time-course of the drugs under study
 - Intervals between test conditions are typically 1 to several days
 - 2) Conducted in a controlled clinical pharmacology laboratory setting
 - Subjects are monitored to prevent use of other drugs
 - 3) Selection of an appropriate subject population
 - Usually subjects with histories of polydrug abuse including drugs from the same pharmacological class as the novel compound, such that subjects can meaningfully rate drug effects and categorize drug effects relative to drugs of known abuse liability
 - The subject population must be one in which the positive control comparison drug tests unequivocally positive
 - 4) Double-blind, placebo controlled drug administration
 - To further reduce expectancy, subjects can be blinded to the specific comparison compound(s) and number of active doses to be administered
 - 5) Selection of appropriate positive control comparison drug(s)
 - The positive control is usually an abused drug from the same pharmacological class and used for the same medical indication proposed for the novel compound
 - Consideration may be given to including a negative control comparison compound from the same class which is behaviorally active but not abused
 - 6) Selection of an appropriate range of doses of the positive control comparison drug
 - Demonstration of orderly dose–effects of the comparison compound on primary outcome measures establishes sensitivity and validity of the trial
 - 7) Selection of a range of doses of the novel compound, including high suprathreshold doses
 - Evaluation of high doses is critical to the validity of the trial
 - A dose run-up pilot study is useful for selecting maximal doses and matching doses between the novel and comparison compound
 - If possible, a high dose of the novel compound should produce effects comparable to the highest dose of the positive control comparison compound on one or more outcome measure
 - 8) Selection of appropriate outcome measures
 - Measures should be assessed repeatedly to characterize onset, peak and duration/offset of drug effects
 - Multiple measures should be included that reflect likelihood of abuse (e.g. liking, good effects, estimated monetary street value, drug vs. money choice)
 - Measures of drug identification and subject-rated side effects and mood changes should be considered in interpreting likelihood of abuse
 - Additional concurrent measures should be included to assess the equivalency of the novel and comparison compounds on some relevant dimensions of biological activity (e.g. behavioral performance, observer-rated assessments, physiological measures)
-

measures related to or reflective of abuse liability. Most of the features listed in the table have been explicitly or implicitly endorsed in reviews on abuse liability assessment of various drug classes (Roache and Griffiths, 1989; Jasinski and Henningfield, 1989; Evans et al., 1991; Bigelow, 1991; Klein et al., 1999), in a 1990 consensus statement about abuse liability testing from an international group of experts (Camí et al., 1991a), as well as in draft guidelines on abuse liability assessment distributed by the Food and Drug Administration (FDA, 1998). This paper involves a synthesis and updating of these previous discussions, providing background and rationale as appropriate.

Before discussing specific features of the acute dose–effect abuse liability trial, it is important to clarify what such a trial can and cannot be expected to accomplish. A dose–effect abuse liability trial with a novel unmarketed compound is most appropriate for predicting the likelihood of abuse by recreational drug abusers and, in turn, the extent of drug diversion and illicit street sales if the novel compound became widely available to the drug abuse community. It is less clear whether these same procedures and data would predict the likelihood of overuse or misuse in patient populations exposed to the drug therapeutically, though many believe these same methods have predictive applicability both to drug abusers and to therapeutically exposed patients. Experimental methods in patient populations have not been developed and validated for predicting the risk of overuse or misuse of medically prescribed drugs in patient populations, with the possible exception of recent studies of nicotine replacement products in cigarette smokers (Schuh et al., 1997; West et al., 2000; Houtsmuller et al., 2002).

2. Typical trial design characteristics

The great majority of dose–effect abuse liability evaluations have used similar experimental designs. Typically, a double-blind within-subject or complete crossover design is used in which all subjects are tested with all drug conditions, although there have been occasional studies that used between group designs (e.g. Preston et al., 1987; Funderburk et al., 1988) and incomplete crossover designs (Jaffe et al., 1983). The drug conditions typically involve placebo, three doses of the novel compound and three doses of a comparison compound. The sequence of conditions is usually mixed (e.g. Latin Square) or randomized. The sample size is typically 10–14 subjects with histories of drug abuse. Each experimental session involves a single administration of drug, which is repeatedly evaluated on the outcome measures over a period of several minutes to several hours, depending on the time-course of the drugs under evaluation. Intervals between experimental ses-

sions are typically 1 to several days, depending on the elimination rates of the drugs under evaluation.

3. Controlled clinical pharmacology laboratory setting

Drug abuse liability evaluations should be conducted in controlled laboratory settings which permit careful assessment of outcome measures in the context of appropriate medical support and prevention of other drug use. For studies with subjects with histories of drug abuse, it is particularly important that volunteers be closely monitored for use of other drugs. Ideally, and most often, such studies are conducted on a closed residential drug abuse pharmacology research unit which provides a stable day-to-day routine and minimizes opportunities for obtaining drugs of abuse. Although such residential studies can also be conducted in other hospital settings such as a General Clinical Research Center, lack of staff training in management of these subjects and lack of structured visiting and inspection policies can be problematic. In some circumstances, studies can be conducted on an outpatient or ambulatory basis, with subjects reporting to the clinical pharmacology laboratory and providing a drug-free urine sample, which is tested prior to drug administration for each session (e.g. Evans et al., 1994; Mumford et al., 1995c; Schuh et al., 2000). Because of drug carryover effects, such studies may require that subjects stay overnight following each drug administration. There are several drawbacks to conducting such trials on an outpatient basis. Subjects with substantial histories of drug abuse can be quite unreliable and often have unhealthy and erratic lifestyles, possibly leading to greater dropout rates and missed visits as well as increased rates of adverse events due to inter-current illnesses, accidents and drug abuse. Also, because sleep, nutrition and drug carryover effects from recent drug abuse are not controlled in outpatient studies, increased variability in outcome measures may occur, with sample size needing to be increased accordingly.

4. Selection of an appropriate subject population

Drug abuse liability evaluations are best conducted in subjects with extensive histories of polydrug abuse, including abuse of drugs from the same pharmacological class as the novel compound, if possible. There are several reasons for this. First, subjects with histories of polydrug abuse can use their prior drug abuse experiences with a variety of drugs and drug classes as a context from which to provide meaningful ratings of the drug experiences in the laboratory. Examples of measures that benefit from extensive histories of drug abuse include ratings of drug high, estimated value on the

street, and categorization of overall drug effect as being similar to specific classes of known drugs of abuse. Second, subjects with histories of drug abuse provide the most face valid population for abuse liability assessment because they represent the population at greatest risk for illicit abuse of a novel compound. Third, in abuse liability studies, subjects with histories of drug abuse generally provide extremely low false positive rates as measured by response to placebo administration. Fourth, in abuse liability studies, subjects with histories of drug abuse are less likely to provide false negative results than subjects without histories of drug abuse. For example, studies with abused barbiturates, benzodiazepines and opioids have shown that normal subjects without histories of drugs abuse generally do not report liking or show self-administration of these compounds (Lasagna et al., 1955; Griffiths and Weerts, 1997). In contrast, subjects with histories of drug abuse show reliable increases in reports of liking and readily self-administer these same drugs (Lasagna et al., 1955; Jasinski, 1977; de Wit and Griffiths, 1991; Griffiths and Weerts, 1997). Fifth, and finally, drug abusers are the actual criterion population whose behavior laboratory studies seek to predict. Vulnerability to drug abuse and to the reinforcing effects of drugs is not spread equally through the entire population. Whatever the combination of genetic, neurobiological, and psychosocial factors that contribute to this unequal distribution of vulnerability, experienced drug abusers have demonstrated that they are in the vulnerable group.

The subject population selected for an abuse liability evaluation must be one in which the positive control comparison drug will test unequivocally positive. The most unambiguous results can be obtained from studies with subjects who have extensive histories of polydrug abuse. If such subjects are not available, it may be possible to use subjects with less extensive drug abuse histories. One approach that has been used for identifying appropriate subjects for abuse liability studies who do not have extensive polydrug abuse has been to pre-screen volunteers with a test dose of a standard abused drug (e.g. Busto et al., 1994, 1999; Zawertailo et al., 1995; Baylon et al., 2000). Volunteers who report liking the standard compound can then be admitted into the abuse liability trial.

Although it is most prudent to initially test any novel drug for abuse liability in subjects with histories of drug abuse, a case can be made that the abuse liability of a novel amphetamine-like stimulant could be reliably assessed in normal subjects. This is a drug class for which vulnerability is very widespread. Studies have shown that subjects without histories of drug abuse report euphoria and liking and show self-administration of amphetamine-like drugs (Lasagna et al., 1955; Foltin and Fischman, 1991a), although there may be considerable individual differences (de Wit et al., 1986). How-

ever, abuse liability assessment studies with non-abusers may be difficult to interpret. Consider the possible outcomes of such a study conducted in non-abusers, which evaluated a novel stimulant and used amphetamine as a positive control. Positive results from such a study would suggest, probably correctly, that the novel compound did indeed have amphetamine-like abuse liability. However, negative results would be inconclusive, especially if the novel drug had some actions that were not amphetamine-like; the possibility would remain that the novel drug would have tested positive in drug abusers, who are known to have lower false-negative rates with other drug classes, as described previously.

Likewise, a case could be made that the abuse liability of a novel anxiolytic could be validly assessed in subjects with histories of moderate alcohol consumption but without histories of drug abuse per se. Studies have shown that diazepam is liked and serves as a reinforcer in subjects with histories of moderate alcohol consumption, but not in subjects with histories of only light social alcohol consumption (de Wit et al., 1989; de Wit and Doty, 1994; Evans et al., 1996). Thus, it would be possible to conduct an abuse liability evaluation of a novel anxiolytic and conclude that it had diazepam-like abuse liability. However, analogous to the discussion of the novel stimulant above, negative findings in such a study would be inconclusive due to the limited sensitivity of the study population.

Finally, a case could be made for evaluating a novel nicotine formulation being developed for treatment of nicotine dependence in subjects with histories of cigarette smoking but without histories of drug abuse. In fact, such studies have been conducted (Schuh et al., 1997; West et al., 2000; Houtsmuller et al., 2002).

In conclusion, the subject population selected for an abuse liability evaluation must be one in which the positive control comparison drug will test unequivocally positive. The most unambiguous results can be obtained from studies with subjects who have extensive histories of polydrug abuse.

5. Double-blind, placebo-controlled drug administration

Standard clinical pharmacology methods require that the subject and the staff who interact with the subject must remain uninformed about which specific drug conditions administered on a given session. In addition, a “no-effect” control in the form of placebo is standard in abuse liability evaluations as a further protection from expectancy or accidental bias influences. Because expectancy can play such an important role in any measures of subjective effects, it is also desirable to keep subjects and staff blind to the specific drugs and drug conditions under study. In a study of a novel compound

for which limited human clinical data are available, it would likely be an ethical requirement to identify the novel compound and its known toxicity profile. However, subjects do not need be informed of how many active doses of the test compounds will be administered or the specific comparison compound(s) that might be administered. Blinding can be enhanced, and the risk of expectancy biases can be minimized, by providing subjects with more information than is applicable to that specific study. For example, subjects can be informed that the study will involve administration of one or more doses of the specific novel drug and, in addition, may involve administration of placebo and a wide range of mood altering drugs. The risks listed in the consent form can include the side effects of drugs not being administered. In this type of optimally blinded study, the consent form lists a series of other drug classes, as well as specific examples of approved medical drugs from these classes (e.g. 1. Sedatives and anti-anxiety drugs such as diazepam (VALIUM®), alprazolam (XANAX®), buspirone (BUSPAR®), pentobarbital, zolpidem (AMBIEN®); 2. Stimulants and weight loss medications such as such as D-amphetamine (DEXEDRINE®), caffeine (such as NO-DOZ®, VIVARIN®), and diethylpropion (TENUATE®); 3. Antihistamines, such as diphenhydramine (BENADRYL®), and chlorpheniramine (such as RYNATAN®); 4. Opioids such as morphine, methadone (DOLOPHINE®), butorphanol (STADOL®); 5. Neuroleptics such as chlorpromazine (THORAZINE®), haloperidol (HALDOL®); and 6. antidepressants such as citalopram (CELEXA®) trazodone (DESYREL®), fluoxetine (PROZAC®)).

6. Selection of appropriate positive control comparison drug(s)

A well-established principle of any biological assay is that the unknown sample must be compared with a standard compound tested under identical assay conditions. Use of a pharmacologically inactive negative control condition alone (i.e. placebo) does not verify the sensitivity of the assay. The biobehavioral assay of abuse liability assessment involves comparing the novel compound to a standard compound of known abuse liability (i.e. a positive control). The positive control should have measurable abuse liability previously established through experimental studies and epidemiological data. If possible, the positive control should be an abused drug from the same pharmacological class and used for the same medical indication as that proposed for the novel compound. For example, although morphine and amphetamine are readily abused, it probably would not be meaningful to use these compounds as positive controls in testing a novel benzodiazepine-like

anxiolytic drug. However, if the novel anxiolytic had a strong dopaminergic mechanisms of action, it might be reasonable to consider amphetamine as a positive control. Because onset and duration of action can affect abuse liability, interpretation of the results of an abuse liability evaluation will be facilitated if the positive control and novel compound have similar onsets and durations. A final consideration in selecting an appropriate positive control compound is the pharmacokinetic elimination rate. Because abuse liability trials commonly use within-subject experimental designs in residential laboratory settings, selection of a positive control comparison compound that is quickly eliminated permits scheduling sessions at more frequent intervals, thus increasing the efficiency of the study. Thus in evaluations of anxiolytics, shorter-acting compounds such as lorazepam or alprazolam would be better positive control choices than would the slowly eliminated diazepam. Similarly, in evaluation of hypnotics, triazolam would be a better choice than flunitrazepam; and for evaluating opioids morphine or hydromorphone would be a better choice than the longer-acting methadone. In addition to a positive control comparison drug, for the purposes of evaluating some novel compounds it may be desirable to consider adding a pharmacologically active negative control compound. Such a compound would be expected to produce significant subjective ratings without producing substantial elevations in measures reflecting likelihood of abuse. Unfortunately, relatively few studies have been conducted using such pharmacologically active negative control compounds in abuse liability assays. Examples of such compounds include caffeine as a stimulant (e.g. Chait and Griffiths, 1983), diphenhydramine as a hypnotic (e.g. Preston et al., 1992), and pentazocine or butorphanol as an opioid (Preston et al., 1987, 1989; Preston and Bigelow, 1994).

7. Selection of an appropriate range of doses of the positive control comparison drug

Ideally, three or four dose levels of the positive control compound are included, spanning a 3–8 fold range of doses, and producing effects ranging from little to moderate or high. Within the study, the positive control compound should demonstrate dose-related statistically significant increases on the primary measures of abuse liability. Failure to demonstrate significant increases with the positive control drug invalidates the assay. One purpose of testing a range of doses is to determine the limits of sensitivity of the assay to lower doses. For example, suppose an abuse liability assessment with a novel anxiolytic were conducted in which the novel compound was compared only to a single high dose of a standard anxiolytic (e.g. 4 mg alprazolam).

Even if the standard compound significantly increased measures of likelihood of abuse and the novel compound did not, it could not be concluded with certainty that the novel compound was without abuse liability because the sensitivity of the assay to lower doses of alprazolam (e.g. 2 mg), which are known to have abuse liability, was not assessed. As discussed in more detail below, a final reason for testing a range of doses is that this permits comparison of the slopes of the dose–effect functions across different measures, which may be crucial to concluding that the novel compound has less abuse liability.

8. Selection of an appropriate range of doses of the novel compound, including high supratherapeutic doses

It is essential to the validity of an abuse liability trial that a sufficiently high supratherapeutic dose of the novel compound be tested. The principal purpose of the initial abuse liability evaluation is to determine whether the compound is likely to lead to public health and social problems because of its abuse by drug abusers. With any new drug, it must be assumed that drug abusers will inevitably sample high doses and the effects of these doses will determine whether or not the drug engenders problematic patterns of continuing use and abuse. In this regard, it should also be recognized that communication among drug abusers about new drugs of abuse has been greatly facilitated by the Internet, with currently active internet sites providing detailed information about dosing and availability of compounds with abuse potential. There can be no safe assumption that potential drug abusers will remain uninformed about the dose levels or others steps needed to make drug use reinforcing. Another reason for testing high doses is that drug abusers tend to be remarkably insensitive to drugs of abuse, possibly reflecting behavioral or physiological tolerance development and/or a macho disposition to significantly under-report magnitude of subjective drug effects.

It should be clear that, in determining the highest dose of the novel compound to be tested, the planned therapeutic dose of the compound is only very marginally relevant. Recreational drug abusers will not be guided by the Patient Package Insert in their selection of doses. Furthermore, consumption of a large number of pills or capsules (e.g. even 10–20) is not a meaningful deterrent for drug abusers. Perhaps the only relevance of the therapeutic dose would be if the therapeutic dose were so many times lower than the abused dose that abuse would be practically or economically unlikely. Table 2 illustrates the lack of meaningful relationship between the therapeutic dose and the highest dose evaluated in an abuse liability trial by summarizing information from various sedative drugs that were

Table 2
Relationship between the therapeutic dose and the high dose evaluated in an acute dose–effect abuse liability trial for various sedatives

Drug	Usual high recommended acute therapeutic dose ^a (mg)	High dose in abuse liability trial (mg)	Ratio of high dose in abuse liability trial to therapeutic dose	Reference
Triazolam (Halcion [®]) hypnotic	0.25 ^b	3.0	12	Roache and Griffiths, 1985
Pentobarbital (Nembutal [®]) hypnotic	100	600	6	Roache and Griffiths, 1985
Meprobamate (Miltown [®]) anxiolytic	600	3600	6	Roache and Griffiths, 1987
Lorazepam (Ativan [®]) anxiolytic and hypnotic	4	9	2.25	Roache and Griffiths, 1987
Methocarbamol (Robaxin [®]) muscle relaxant	500	12 000 ^c	24	Preston et al., 1989
Zolpidem (Ambien [®]) hypnotic	10	80 ^c	8	Evans et al., 1990
Diphenhydramine (Benadryl [®]) hypnotic	50	600 ^c	12	Preston et al., 1992
Buspirone (BuSpar [®]) anxiolytic	7.5	120	16	Troisi et al., 1993
Tandospirone (Sediel [®] , Japan) anxiolytic	10	160	16	Evans et al., 1994
Alprazolam (Xanax [®]) anxiolytic	0.5	4.0	8	Mumford et al., 1995a
Abecarnil (not marketed) anxiolytic	7.5	100 ^c	13.3	Mumford et al., 1995a
Pazinaclone (not marketed) anxiolytic	4	32	8 ^d	Mumford et al., 1995b
Flunitrazepam (Rohypnol [®]) hypnotic	1 ^b	8	8	Mintzer and Griffiths, 1998
Zaleplon (Sonata [®]) hypnotic	10	75	7.5	Rush et al., 1999

^a Usual high initial acute therapeutic dose based on dosing recommendations published in the Physicians' Desk Reference[®] or based on published literature.

^b Higher dose formulations were withdrawn because of adverse effects or abuse.

^c Based on the published dose run-up study from which doses were selected for the subsequent comparative dose–effect trial.

^d Pazinaclone (DN-2327) was withdrawn from development, in part, because of strong sedative effects relative to expected anxiolytic effects.

evaluated in acute dose–effect abuse liability trials. The median ratio of high dose used in the abuse liability trial to the therapeutic dose was 8:1, with ratios ranging from 2.25:1 to 24:1. It is instructive to review the data from the trials summarized in Table 2 to consider what conclusions might have been reached had the high dose in the abuse liability trials been arbitrarily constrained to only two to three times the high therapeutic dose. For over half the drugs evaluated, the data would have suggested that these compounds were either inactive, or produced only weak behavioral and subjective effects. For example, the trial would have totally missed the fact that buspirone produces dose-related increases in disliking and dysphoric symptoms and has negligible abuse liability. A similar outcome suggesting zaleplon to be inactive would have missed the conclusion that zaleplon produces dose-related increases in liking, good effects and street value and has classic benzodiazepine-like abuse liability. Overall, it is clear that there is no meaningful rule of thumb relating therapeutic dose to the appropriate selection of the highest dose for abuse liability assessment of sedative drugs. We do not know whether a similar lack of relationship exists for other drug classes. A prudent course in evaluating any pharmacologically novel compound would be to select the high dose based on data from pilot studies (as described below) rather than on proposed therapeutic dose levels.

If possible, the high dose of the novel compound should be selected that will produce effects comparable to the highest dose of the positive control comparison

compound on one or more outcome measure. Ideally, a range of lower doses of the novel drug should also be tested such that it is possible to compare the dose–effect curves for the novel and standard compounds across a range of different measures. No standard guidelines are possible for selecting dose increments because the slopes of the dose–effect functions and the high dose toxicology profiles may differ markedly across different compounds.

Understandably, clinical investigators, industry sponsors, Institutional Review Boards and the Food and Drug Administration prefer to be conservative about approving the use of high doses for which limited safety data are available. Hesitancy about testing high doses also reflects the concern by industry sponsors that new untoward effects might be observed that would slow or even jeopardize final approval of the drug for marketing. Nonetheless, an adequate dose range is integral to a meaningful abuse liability evaluation and merits the extra precautions that may be required to gain approval for the protocol. If a drug has problems, it is better to find them early rather than fail to look.

Fig. 1 illustrates an unambiguous outcome in an abuse liability trial that evaluated an appropriate range of doses of a novel compound and a positive control compound. The purpose of this trial (Mumford et al., 1995b) was to evaluate the relative abuse liability of the novel β -carboline anxiolytic, abecarnil. Alprazolam, an anxiolytic benzodiazepine which is known to be abused, was selected as the positive control compound. Doses of abecarnil and alprazolam were selected on the basis of a

dose run-up pilot study. An analysis of time-course of subjective and performance measures showed that the two drugs produced similar dose- and time-related effects, with roughly comparable time to onset, peak and offset of effects. The left column of Fig. 1 shows that alprazolam and abecarnil produced parallel dose effect-curves and similar maximal effects on three measures of behavioral and cognitive performance. Similar comparable effects were shown on subject- and staff-ratings of strength of drug effect. These data indicate that comparable doses of alprazolam and abecarnil were selected for study. The right column of Fig. 1 shows non-parallel dose-effect curves and clear differences between the drugs on several measures of

abuse liability. Alprazolam produced significant dose-related increases in these measures in contrast to abecarnil which did not. Furthermore, alprazolam significantly increased a drug versus money choice measure of reinforcement while abecarnil did not (data not shown). Finally, the high dose of abecarnil produced significant elevations in subject ratings of “bad” effects and “disliking,” which were not significantly affected by alprazolam. The highest dose of abecarnil evaluated in this study (40 mg) was considerably higher than the highest acute anxiolytic doses evaluated in therapeutic trials (3 and 7.5 mg). The overall conclusion, that abecarnil does not have alprazolam-like abuse liability, was consistent with preclinical data which showed that

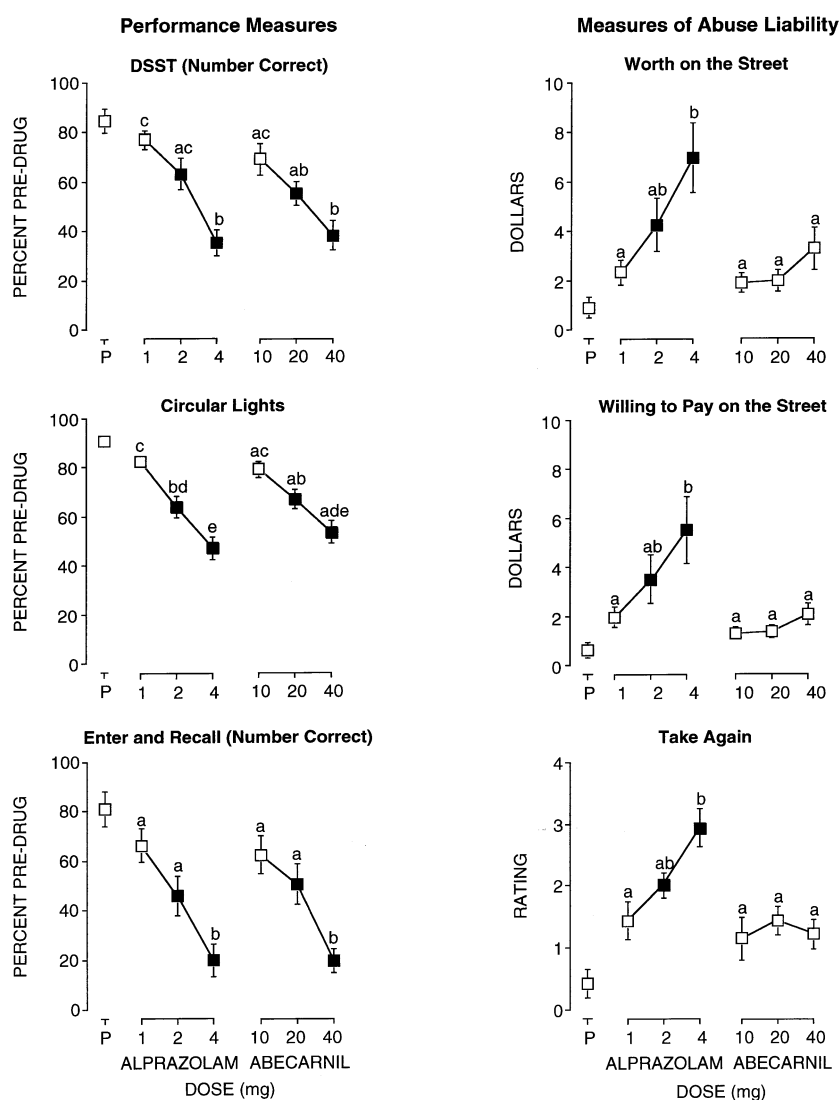


Fig. 1. Dose effects for the peak magnitude of performance impairment (left column) and several measures of abuse liability from the Next Day Questionnaire (right column). Peak effect data for the DSST (a measure of cognitive performance and speed), circular lights (a measure of psychomotor performance speed) and enter and recall (a measure of short-term memory) are expressed as percentages of pre-drug scores. X-axes: dose in mg (log scale); data points at “P” designate placebo values. Data points show means of 14 subjects; vertical bars show ± 1 S.E.M. and the absence of bars indicates 1 S.E.M. fell within the area of the data symbol. Filled symbols indicate a significant difference from placebo ($P \leq 0.05$). Letters “a”, “b”, “c”, “d” and “e” indicate comparisons among the six drug doses; within the same panel any two means designated with the same letter are not significantly different from each other at $P \leq 0.05$ (Tukey’s post hoc tests) [adapted from Mumford et al., 1995b].

abecarnil does not maintain self-injection in non-human primates (Sannerud et al., 1992).

8.1. Dose run-up pilot study

A useful approach for selecting maximal doses of the novel compound and for matching doses between the novel compound and the comparison compound is to conduct a dose run-up pilot study in subjects with histories of drug abuse. Often the previous Phase I or Phase II studies conducted by the sponsor provide little guidance for dose selection because they typically use a limited range of doses in drug inexperienced subjects. Also, these previous studies often focus on a narrow range of outcome measures such as anticipated therapeutic effects or specific toxicities. A dose run-up pilot study in a drug abuser subject population can provide the needed empirical base for dose selection. Much can be learned by exposing a few (4–8) subjects to a full range of doses of the novel compound plus placebo and one or two doses of the comparison compound. An ascending order of doses of the novel compound maximizes safety by permitting termination of the dose escalation in subjects showing side effects that likely would not be tolerated or would be medically dangerous in the subsequent mixed- or random-sequence dose-effect abuse liability trial. Because the dose run-up may be considered a Phase I safety study, it may be necessary to include physiological, blood chemistry and other safety measures to establish the feasibility and safety of using high doses in the subsequent dose-effect study. The dose run-up also provides an opportunity to evaluate and modify the measures and procedures for the subsequent dose-effect comparison study. For example, the adequacy of the washout duration for eliminating possible pharmacodynamic carryover drug effects from previous sessions can be evaluated. Importantly, selection and refinement of measures to define comparable effects between the novel and standard compounds can occur.

9. Selection of appropriate outcome measures

Three types of outcome measures should be considered in dose-effect abuse liability trials: 1. measures most directly related to predicting the likelihood of abuse (e.g. ratings of liking, disposition to take the drug again, street value and drug vs. money choice behavior); 2. measures that should be considered when interpreting likelihood of abuse (e.g. drug identification, subject-rated side effects and mood changes); and 3. other concurrent measures of drug effect (e.g. subject-rated strength of drug effect, behavioral and cognitive performance, observer-rated measures, physiological effects). Ideally, the assessment procedures should be easily

learned and rapidly completed by subjects, and show good test-retest reliability so that they can be collected repeatedly within a session to track the time course of drug effects. Subjects should be familiarized with all measures before initiating the trial. With performance tasks, subjects should acquire sufficient proficiency to minimize continued acquisition of performance skill during the trial. With questionnaires, subjects should be familiarized with the vocabulary and specific questions. It is important to assess the time-course of effects by repeatedly assessing outcome measures. The spacing of assessment intervals should be sufficient to characterize onset, peak, and offset of effects. For example, with an orally active hypnotic or analgesic outcome measures might be assessed at 0.5, 1, 2, 3, 4, 5, 6, 8, 12 and 24 h after drug administration (Rush et al., 1999). For a short-acting intravenously administered stimulant, in contrast, assessments might occur at 1 min intervals for 20–30 min after drug administration (e.g. Jones et al., 1999; Abreu et al., 2001).

9.1. Measures most directly related to predicting likelihood of abuse

9.1.1. Subject-ratings of likelihood of abuse

Measures of drug liking (i.e. asking subjects how much they “like” the drug) have face validity, have been used in most studies of abuse liability, and tend to be one of the most sensitive and reliable measures of likelihood of abuse. Liking scales can be either unipolar (ranging from “neutral” or “not at all” to “like very much”) or bipolar (ranging from “dislike very much” to “like very much”). If a unipolar scale is used, there is a risk that the liking will reflect only those effects subjects consider positive and will ignore concurrent unpleasant effects. Although not empirically demonstrated, it is possible that subjects are more likely to provide their overall global evaluation of liking with a bipolar scale. Other subject ratings that generally co-vary with liking include ratings of “good effects”, “bad effects”, “degree to which you would like to take the drug again”, “estimate of street value of the drug”, and “estimate of amount of money you would personally be willing to pay for the drug.” In addition to obtaining time-course data, these measures can also be assessed after the drug effects have dissipated (e.g. a Next Day questionnaire, given 24 h after drug administration). These retrospective ratings have the advantage of assessing the overall drug experience, or at least the remembered portion of that experience, under drug-free conditions and thus may provide valuable indices of the likelihood that an individual, when sober and drug-free, would seek out an opportunity to re-administer the compound previously sampled.

9.1.2. *Drug reinforcement assessed with the drug versus money Multiple Choice Procedure*

A relatively new innovation in dose–effect abuse liability evaluations has been the addition of a behavioral measure of drug versus money choice to provide a complementary assessment of likelihood of abuse. The Multiple-Choice Procedure was developed and validated as a tool to efficiently assess drug reinforcement in humans (e.g. Griffiths et al., 1993, 1996). The drug versus money version of the procedure has been adapted to provide a contingency-based assessment of the monetary value of each drug condition tested in dose–effect comparison abuse liability evaluations (e.g. Mumford et al., 1995a,b,c; Mintzer and Griffiths, 1998; Jones et al., 1999). With this procedure, after each session the subject makes a series of choices between receiving various amounts of money or receiving the drug condition one more time in a final session at the end of the protocol. The data from the procedure is fundamentally different from simple subject ratings of liking or estimated street value because the choices have a tangible consequence to the subject in terms of money earned or drug received; for the final study session, one of the subject's choices is randomly selected and the subject actually receives the consequence of that choice. The Multiple-Choice Procedure has been studied in the context of evaluating sedatives (Mumford et al., 1995a,b,c; Mintzer and Griffiths, 1998; Haurert et al., 2002) and stimulants (Jones et al., 1999; Schuh et al., 2000; Roache et al., 2002; Smith et al., 2001; Jones and Griffiths, 2003), but it has not been used to evaluate drugs from other pharmacological classes. A variation of the procedure permits it to also assess the negative reinforcing effects of a drug (i.e. how much money a subject is willing to forfeit so as not to receive the drug again) (Schuh and Griffiths, 1997; Garrett and Griffiths, 1998; Schuh et al., 2000; Jones and Griffiths, 2003). Although data from the Multiple-Choice Procedure and from drug self-administration procedures usually covary with subject ratings of likelihood of abuse, dissociation between subjective effects and drug reinforcement have been noted in previous studies (e.g. Lamb et al., 1991; Jones and Griffiths, 2003). Given that the Multiple-Choice Procedure represents a relatively recent innovation in abuse liability assessment, it is premature to recommend it as the primary outcome measure for a pivotal abuse liability trial, however it deserves consideration as one of the secondary outcome measures.

9.2. *Measures that should be considered in interpreting likelihood of abuse*

Although only indirectly related to likelihood of abuse, there are at least two types of measures that can be useful in interpreting and explaining the measures

of likelihood of abuse: drug identification, subject-rated side effects and mood changes.

9.2.1. *Drug identification*

In many abuse liability studies, subjects with extensive histories of abusing a range of other drugs are asked to categorize the effect of the test drug as being most similar to one of as many as 14 different classes of psychoactive drugs. The questionnaire provides descriptive titles for, and examples of, each of the classes (e.g. blank or placebo, opiate, antipsychotic, muscle relaxant, barbiturate and sleeping drug, hallucinogen, benzodiazepine, stimulant, alcohol, cocaine, marijuana, phencyclidine, antidepressant and other). The questionnaire provides a useful estimate of the drug class with which drug abusers identify the novel drug. The relationship between drug identification and drug liking (or reinforcement) in humans is analogous to the relationship between drug discrimination and drug reinforcement in preclinical studies (cf. Ator and Griffiths, 2003, this volume). Drug identification or drug discrimination alone does not provide a meaningful estimate of abuse liability. For example, in the study described in Fig. 1, which compared abecarnil and alprazolam (Mumford et al., 1995b), the high doses of both drugs were identified by most subjects (about 85%) as being either a benzodiazepine or barbiturate, yet the drugs showed clear differences in abuse liability. On the other hand, in a different study with a novel sedative (Rush et al., 1999), the high doses of the novel hypnotic (zaleplon) and the comparison compound (triazolam) both produced high rates of identification as a benzodiazepine or barbiturate and both compounds showed similar increases in liking and other measures of likelihood of abuse. In these cases, it could be concluded that zaleplon had benzodiazepine-like abuse liability but that abecarnil did not.

9.2.2. *Subject-rated side effects and mood changes*

The acute dose–effect comparison study provides an excellent opportunity to systematically assess side effects and mood changes, which may have relevance to abuse liability. For sedative and anxiolytic drugs, for example, sensitive subject-rated side effects have included ratings of blurred vision, confusion, difficulty concentrating, drug strength, easy going, forgetful, heavy limbs, limp/loose, mentally slowed down, relaxed, sedated, sleepy, slurred speech, shaky/jittery and numb/tingling (e.g. Evans et al., 1990; Preston et al., 1992; Rush, et al., 1999). Obviously, different side effects should be assessed for different drug classes. With opioids and analgesics, for example, ratings of flushing, skin itchy, sweating, turning of stomach, nodding, relaxed, coasting or spaced out, pleasant sick, talkative, heavy or sluggish feeling appear to provide good sensitivity to opioid agonist effects (cf. Bigelow, 1991). For stimulants, in contrast, sensitive subject-rated adjectives have included

stimulated, shaky or jittery, nervous or anxious, and irregular or racing heart (cf. Jones et al., 1999; Rush et al., 2002). Often adjective ratings are grouped and scored as “scales” reflective of the symptom clusters commonly associated with prototypic effects of particular drug classes (e.g. sedation scale or an opioid agonist scale, Bigelow, 1991).

9.3. Other concurrent measures of drug effect

Related to the previous discussion about selection of comparable doses of the novel and comparison compound, it is important to assess a variety of measures of the effects of the study drugs. By constructing dose–effect functions across a range of measures, it becomes possible to determine doses of the novel compound and the comparison compound that produce similar effects on some measures. Conclusions about differences in likelihood of abuse are most meaningful in the context of empirically demonstrated dose comparability on some measures. For example, in the previously discussed study comparing abecarnil and alprazolam (Fig. 1), the finding that alprazolam produced greater elevations in measures of likelihood of abuse than did abecarnil was especially meaningful given that both drugs produced comparable dose-related performance impairment and ratings of strength of drug effect.

Evaluation of a variety of concurrent measures is also desirable to the extent that the novel compound may show a unique profile of effects. The acute dose–effect comparison study provides an excellent opportunity to characterize carefully a range of pharmacological differences between the novel and standard compound, particularly those that might be relevant to clinical use. Several types of concurrent measures can be assessed: subject-ratings of strength of drug effect, behavioral and cognitive performance, observer-rated measures, physiological measures.

9.3.1. Subject-rated strength of drug effect

Most studies of abuse liability include a subject rating of the magnitude of the strength of drug effect on a unipolar scale (ranging, for example, from “not at all” to “very strong drug effect” or “extremely”). This rating provides critical information about the subjective detectability of any drug effect. When testing and comparing drugs that produce detectable effects, it is best to test doses of the novel drug and comparison compound that show similar dose-related increases in strength of drug effect, suggesting that the discriminability of the two drugs was equivalent at the doses tested. It is important to note, however, that subject ratings of drug strength do not always co-vary with other measures of drug effect. For example, studies with several benzodiazepines show that subjects tend to under-rate magnitude of drug effect relative to staff ratings and performance

impairment (Roache and Griffiths, 1985, 1987). Also, studies with buspirone and tandospirone, in contrast, show that these drugs produce robust increases in subject ratings of drug strength at doses that have no or minimal effects on behavioral and cognitive performance (Troisi et al., 1993; Evans et al., 1994).

9.3.2. Behavioral and cognitive performance

In principle, sensitive concurrent measures of behavioral and cognitive performance could be useful in abuse liability evaluations of any type of novel compound. In practice, such measures have been most widely used in abuse liability assessments of anxiolytics and hypnotics because behavioral and cognitive impairments represent well-recognized adverse side effects of these drugs. Examples of sensitive behavioral and cognitive measures include balance, hand–eye psychomotor speed and coordination (e.g. Circular lights task), reaction time, cognitive performance (e.g. Digit Symbol Substitution task), short-term or working memory (e.g. Enter and recall task), and long-term memory (e.g. Picture recall/recognition task) (cf. Evans et al., 1990; Rush et al., 1999). Clearly, a different battery of concurrent behavioral and cognitive measures may be appropriate for different drug classes. With opioids little behavioral performance impairment is typically seen (Hill and Zacny, 2000).

9.3.3. Observer-rated measures

Observer-rated measures of drug effect have been included in most abuse liability studies, with the nature of the measures tailored to the expected effects from the class of drug being investigated. For anxiolytics and hypnotics, sensitive observer-rated measures have included assessments of sleep, sedation, muscle-relaxation, impaired posture, impaired speech, confusion, and overall strength of drug effect (cf. Evans et al., 1990; Rush et al., 1999). With opioids and analgesics, sensitive observer ratings have included nodding, carefree, talkative, scratching, relaxed, coasting, vomiting, sleepy, nervous and drunken (Jasinski, 1977; Walsh et al., 1995).

9.3.4. Physiological measures

Sensitive physiological measures are often considered a desirable additional concurrent measure in abuse liability studies because of their presumed complete objectivity. Such measures may have no evident relation to the construct of abuse liability, but they contribute importantly to assessing the pharmacological equivalence of novel compounds to reference compounds of known abuse liability. Concurrent examination of certain physiological measures (e.g. blood pressure, respiratory rate) may be added to abuse liability trials for purposes of safety assessment. With opioids and analgesics, pupillary diameter provides a frequently used

and sensitive physiological measure of opioid action (cf. Bigelow, 1991); it can also be useful in assessing CNS stimulants, which may produce pupillary dilation, in contrast to the constriction characteristic of opioids (Walsh et al., 1996). Other sensitive physiological measures for stimulants include heart rate, blood pressure and skin temperature (cf. Jones et al., 1999; Rush et al., 2002). For sedatives, post-rotatory nystagmus has been shown to be a sensitive dose-related measure (cf. Fraser and Jasinski, 1977); but it is not widely used because the procedure is cumbersome and it can be unpleasant and disruptive to subjects.

9.3.5. *Assessment of drug alteration of mood state using the POMS*

The Profile of Mood States (POMS) questionnaire is a 65-item adjective rating scale that is considered to be a standardized mood state inventory with six standard scales (tension–anxiety, depression–dejection, anger–hostility, vigor, fatigue, and confusion–bewilderment) and a composite scale reflecting total mood disturbance (McNair et al., 1992). Some investigators have used modified versions of the POMS which allow scoring of friendliness, arousal, elation and positive mood (Foltin and Fischman, 1991b). The POMS has been sensitive to sedative and stimulant drugs in subjects who were light drug users or nonusers (de Wit and Griffiths, 1991; Foltin and Fischman, 1991a). However, the POMS has not been widely used in studies with subjects with histories of extensive drug abuse. In stimulant abusers, the POMS was shown to be sensitive to the effects of intravenously administered cocaine (Foltin and Fischman, 1991a). In sedative abusers, the POMS has been relatively insensitive, sometimes failing to show robust dose-related effects that are clearly shown on other measures (de Wit and Griffiths, 1991).

9.3.6. *The Addiction Research Center Inventory (ARCI)*

The short form of the ARCI is a 49 item true/false questionnaire with five empirically derived scales (MBG, morphine–benzedrine group scale, as an index of euphoria; LSD, lysergic acid diethylamide specific scale as an index of dysphoria and somatic symptoms; PCAG, pentobarbital–chlorpromazine–alcohol group scale as an index of sedation; A, amphetamine and BG, benzedrine group scales are both indices of amphetamine-like effects)(Martin et al., 1971; Jasinski, 1977).¹ The ARCI has been extensively used, both in studies with normal subjects and subjects with histories of drug abuse, to assess the effects of opioids, stimulants and sedatives (Jasinski, 1977; Bigelow, 1991; Foltin and Fischman, 1991a; de Wit and Griffiths, 1991). Studies generally show significant elevations in scales appropriate to the drug class being tested. Of particular relevance to abuse liability assessment trials is the MBG scale, which is often mistakenly thought to

provide a selective measure of drug-induced euphoria for any drug class. The MBG scale was originally derived from a common pattern of items that were rated “true” after administration of morphine or benzedrine (Haertzen, 1966). It consists of items related to pleasant feelings, social effectiveness, contentment, popularity, energy and clarity. Perhaps not surprisingly, scores on the MBG scale have frequently demonstrated dose-related elevations after administration of morphine-like opioid agonists and amphetamine-like stimulants, including cocaine (cf. reviews by Jasinski, 1977; Foltin and Fischman, 1991a). However, this is not the case with sedative drugs. The classic abused sedative, pentobarbital, and the abused benzodiazepines, diazepam and alprazolam, do not reliably increase MBG scores in subjects with histories of drug abuse, although liking and other measures of likelihood of abuse show significant and orderly dose-related increases in these same studies (de Wit and Griffiths, 1991; Evans et al., 1994; Mumford et al., 1995a,c). These observations indicate that the MBG scale is not a valid predictor of the abuse potential of sedative drugs and suggest the possibility that it may not have broad utility in assessment of novel drugs other than those that are selectively morphine- or amphetamine-like. The MBG scale, and the other ARCI scales, are abuse liability assessment tools developed decades ago when the field was just beginning. Although they have played and continue to play a valuable role in assessment of abuse liability, they are sometimes less sensitive than other more recent assessment questionnaires (Bigelow, 1991; Preston and Bigelow, 1994).

10. Limitations and future directions

This paper outlines the features of the current “gold standard” for initial abuse liability assessment of a novel drug—the classic acute dose–effect comparison study. Because research in abuse liability assessment will continue to develop and refine these methods, and because novel pharmacological agents or dosage forms often present special considerations, none of the features outlined in Table 1 and discussed in the preceding sections should be considered as an inflexible mandate. Rather, the features should be viewed as strong recommendations to be carefully considered in the design of an abuse liability trial. By way of illustration, it is readily conceivable that a completely novel pharmacological agent could be developed for which no single comparison compound would be appropriate. Under

¹ Subscales of the ARCI have been translated into Spanish and the validity of the translated versions has been examined. These include the 49-item short form questionnaire (Arasteh et al., 1999).

these circumstances, consideration could be given to conducting two or more trials using two or more comparison compounds. Alternatively, a single trial could be designed involving two or more comparison compounds. For such a trial, it may not be practical to include three dose levels of all compounds. This could result in a loss of the ability to compare dose–effect functions. Under some circumstances this may be judged to be an acceptable trade-off. However, it should be recognized that there are also circumstances in which comparative dose–effect information is crucial to concluding that the novel compound has less abuse liability. Thus, the risks and benefits of deviating from the gold standard features should be carefully weighed.

A further limitation of this paper is that it considers abuse liability testing only from the perspective of acute dose–effect comparison trials. While this approach to abuse liability assessment is currently the best developed, consideration could be given to using other methods to assessing other domains of drug action thought to be relevant to abuse liability such as the development of tolerance or physical dependence, or the assessment of repeated drug self-administration or the ability to abstain from ongoing drug self-administration. Research probing such various aspects of abuse liability has been described elsewhere (cf. [Fischman and Mello, 1989](#); [Camí et al., 1991b](#)) and is beyond the scope of the present review.

Finally, the emphasis of this review on the assessment of abuse liability in polydrug abusers does not preclude the need for the development of methods for predicting the risk of overuse and or misuse of medically prescribed drugs in the therapeutically exposed patient populations. The relatively low rate of nonmedical drug abuse problems arising from legitimate medical practice makes validation of such methods difficult. Perhaps the best examples of the development of such methods are recent studies which have evaluated several forms of nicotine replacement products in cigarette smokers without histories of drug abuse ([Schuh et al., 1997](#); [West et al., 2000](#); [Houtsmuller et al., 2002](#)). Such studies provide important information about possible overuse or misuse of these products in the population to whom the products will be marketed. Although such studies are clearly important, they do not necessarily rule out the need for conducting studies in polydrug abusers to address the risk of possible organized drug diversion and illicit street sales. One could imagine, for example, that certain rapid onset formulations of a nicotine-like substance might have substantial abuse liability in subjects with polydrug or cocaine abuse histories (cf. [Jones et al., 1999](#)), but not in subjects with histories of only cigarette smoking. Thus, the development of methods for assessing drug misuse in therapeutic populations will ultimately provide information com-

plementary to existing methods assessing abuse liability in polydrug abusers.

11. Conclusions

The present paper describes the rationale and procedures for conducting the classic acute dose–effect comparison abuse liability assessment study in volunteers with histories of drug abuse. Such a trial is appropriate for predicting the likelihood of abuse by drug abusers and represents a crucial initial step in assessing the abuse risk of marketing a novel compound. In addition to predicting the likelihood of abuse, such a study may also provide an assessment of some of the adverse consequences of recreational abuse. The initial dose–effect comparison study will also be helpful to establishing parameters for possible subsequent studies using other methodologies. Subsequent studies might include comparison of the novel compound to other standard compounds, further assessment of drug reinforcement through repeated drug self-administration testing, examination of interactions of the novel compound with frequently abused drugs, examination of effects in special populations (e.g. physically dependent subjects), and evaluation of repeated administration of the novel compound to assess tolerance and physiological dependence development.

Acknowledgements

Preparation of this review was supported by the College on Problems of Drug Dependence and, in part, by National Institute on Drug Abuse grants R01 DA03889, R01 DA03890, R01 DA04133, K05 DA00050, and P50 DA05273.

References

- Abreu, M.E., Bigelow, G.E., Fleisher, L., Walsh, S.L., 2001. Effect of intravenous injection speed on responses to cocaine and hydro-morphine in humans. *Psychopharmacology* 154, 76–84.
- Arasteh, K., Poudevida, S., Farré, M., Roset, P.N., Camí, J., 1999. Response patterns of the Spanish version of the 49-item short form of the Addiction Research Center Inventory after the use of sedatives, stimulants, and opioids. *Drug Alcohol Depend.* 55, 117–125.
- Ator, N.A., Griffiths, R.R., 2003. Principles of drug abuse liability assessment in laboratory animals. *Drug Alcohol Depend.* 70 (3), Suppl. S55–S72.
- Baylon, G.J., Kaplan, H.L., Somer, G., Busto, U.E., Sellers, E.M., 2000. Comparative abuse liability of intravenously administered remifentanyl and fentanyl. *J. Clin. Psychopharmacol.* 20, 597–606.
- Bigelow, G.E., 1991. Human drug abuse liability assessment: opioids and analgesics. *Br. J. Addict.* 86, 1615–1628.

- Busto, U., Kaplan, H.L., Zawertailo, L., Sellers, E.M., 1994. Pharmacologic effects and abuse liability of bretazenil, diazepam, and alprazolam in humans. *Clin. Pharmacol. Ther.* 55, 451–464.
- Busto, U., Zawertailo, L., Kaplan, H.L., Sellers, E.M., 1999. Identifying appropriate subjects for abuse liability studies using prestudy pharmacological testing. *Can. J. Clin. Pharmacol.* 6, 103–110.
- Camí, J., Bigelow, G.E., Griffiths, R.R., Drummond, D.C. (Eds.), 1991a. Barcelona meeting on clinical testing of drug abuse liability: consensus statement and recommendations, *Br. J. Addict.* 86, pp. 1527–1528.
- Camí, J., Bigelow, G.E., Griffiths, R.R., Drummond, D.C. (Eds.), 1991b. Special issue: clinical testing of drug abuse liability, *Br. J. Addict.* 86, Number 12.
- Chait, L.D., Griffiths, R.R., 1983. Effects of caffeine on cigarette smoking and subjective response. *Clin. Pharmacol. Ther.* 34, 612–622.
- de Wit, H., Doty, P., 1994. Preference for ethanol and diazepam in light and moderate social drinkers: a within-subjects study. *Psychopharmacology* 115, 529–538.
- de Wit, H., Griffiths, R.R., 1991. Testing the abuse liability of anxiolytics and hypnotic drugs in humans. *Drug Alcohol Depend.* 28, 83–111.
- de Wit, H., Uhlenhuth, E.H., Johanson, C.E., 1986. Individual differences in the reinforcing and subjective effects of amphetamine and diazepam. *Drug Alcohol Depend.* 16, 341–360.
- de Wit, H., Pierri, J., Johanson, C.E., 1989. Reinforcing and subjective effects of diazepam in nondrug-abusing volunteers. *Pharmacol. Biochem. Behav.* 33, 205–213.
- Evans, S.M., Funderburk, F.R., Griffiths, R.R., 1990. Zolpidem and triazolam in humans: behavioral and subjective effects and abuse liability. *J. Pharmacol. Exp. Ther.* 255, 1246–1255.
- Evans, S.M., Critchfield, T.S., Griffiths, R.R., 1991. Abuse liability assessment of anxiolytics/hypnotics: rationale and laboratory lore. *Br. J. Addict.* 86, 1625–1632.
- Evans, S.M., Troisi, J.R., Griffiths, R.R., 1994. Tandoospirone and alprazolam: comparison of behavioral effects and abuse liability in humans. *J. Pharmacol. Exp. Ther.* 271, 683–694.
- Evans, S.M., Griffiths, R.R., de Wit, H., 1996. Preference for diazepam, but not buspirone in moderate drinkers. *Psychopharmacology* 123, 154–163.
- Fischman, M.W., Mello, N.K., 1989. Testing for abuse liability of drugs in humans, National Institute on Drug Abuse Research Monograph No. 92, National Institute on Drug Abuse, Rockville, MD.
- Foltin, R.W., Fischman, M.W., 1991a. Assessment of abuse liability of stimulant drugs in humans: a methodological survey. *Drug Alcohol Depend.* 28, 3–48.
- Foltin, R.W., Fischman, M.W., 1991b. Methods for the assessment of abuse liability of psychomotor stimulants and anorectic agents in humans. *Br. J. Addict.* 86, 1633–1640.
- Food and Drug Administration, 1999. Guidelines for research involving the abuse liability assessment of new drugs, Food and Drug Administration, Rockville, MD.
- Fraser, H.F., Jasinski, D.R., 1977. The assessment of the abuse potential of sedative/hypnotics (depressants) (methods used in animal and man). In: Martin, W.R. (Ed.), *Drug Addiction I. Morphine, Sedative-Hypnotic and Alcohol Dependence*, Handbook of Experimental Pharmacology, vol. 45. Springer, Heidelberg, pp. 159–196.
- Funderburk, F.R., Griffiths, R.R., McLeod, D.R., Bigelow, G.E., Mackenzie, A., Liebson, I.A., Nemeth-Coslett, R., 1988. Relative abuse liability of lorazepam and diazepam: an evaluation in “recreational” drug users. *Drug Alcohol Depend.* 22, 215–222.
- Garrett, B.E., Griffiths, R.R., 1998. Physical dependence increases the relative reinforcing effects of caffeine versus placebo. *Psychopharmacology* 139, 195–202.
- Griffiths, R.R., Weerts, E.M., 1997. Benzodiazepine self-administration in humans and laboratory animals—implications for problems of long-term use and abuse. *Psychopharmacology* 134, 1–37.
- Griffiths, R.R., Troisi, J.R., II, Silverman, K., Mumford, G.K., 1993. Multiple-choice procedure: an efficient approach for investigating drug reinforcement in humans. *Behav. Pharmacol.* 4, 3–13.
- Griffiths, R.R., Rush, C.R., Puhala, K., 1996. Validation of the multiple-choice procedure for investigating drug reinforcement in humans. *Exp. Clin. Psychopharmacol.* 4, 97–106.
- Haertzen, C.A., 1966. Development of scales based on patterns of drug effects, using the Addiction Research Center Inventory (ARCI). *Physiological Reports* 18, 163–194.
- Hauert, L.A., Helmus, T.C., Lockhart, N., Johanson, C.-E., 2002. Relationship between the multiple-choice procedure, drug choice and subjective ratings. *Drug Alcohol Depend.* 66 (Suppl. 1), S76.
- Hill, J.L., Zacny, J.P., 2000. Comparing the subjective, psychomotor, and physiological effects of intravenous hydromorphone and morphine in healthy volunteers. *Psychopharmacology* 152, 31–39.
- Houtsmuller, E.J., Fant, R.V., Eissenberg, T.E., Henningfield, J.E., Stitzer, M.L., 2002. Flavor improvement does not increase abuse liability of nicotine chewing gum. *Pharmacol. Biochem. Behav.* 72, 559–568.
- Jaffe, J.H., Jaffe, F.K., 1989. Historical perspectives on the use of subjective effect measures in assessing the abuse potential of drugs. In: Fischman, M.W., Mello, N.K. (Eds.), *Testing for Abuse Liability of Drugs*, National Institute on Drug Abuse Research Monograph, No. 92, National Institute on Drug Abuse, Rockville, MD, pp. 43–72.
- Jaffe, J.H., Ciraulo, D.A., Nies, A., Dixon, R.B., Monroe, L.L., 1983. Abuse potential of halazepam and of diazepam in patients recently treated for acute alcohol withdrawal. *Clin. Pharmacol. Ther.* 34, 623–630.
- Jasinski, D.R., 1977. Assessment of the abuse potential of morphine-like drugs (methods used in man). In: Martin, W.R. (Ed.), *Drug Addiction I. Morphine, Sedative-Hypnotic and Alcohol Dependence*. Handbook of Experimental Pharmacology, vol. 45. Springer, Heidelberg, pp. 197–258.
- Jasinski, D.R., Henningfield, J.E., 1989. Human abuse liability assessment by measurement of subjective and physiological effects. In: *Testing for Abuse Liability of Drugs in Humans*, National Institute on Drug Abuse Research Monograph, No. 92, National Institute on Drug Abuse, Rockville, MD.
- Jones, H.E., Griffiths, R.R., 2003. Oral caffeine maintenance potentiates the reinforcing and stimulant subjective effects of intravenous nicotine in cigarette smokers. *Psychopharmacology* 165, 280–290.
- Jones, H.E., Garrett, B.E., Griffiths, R.R., 1999. Subjective and physiological effects of intravenous nicotine and cocaine in cigarette smoking cocaine abusers. *J. Pharmacol. Exp. Ther.* 288, 188–197.
- Klein, M., Calderon, S., Hayes, B., 1999. Abuse liability assessment of neuroprotectants. *Ann. NY Acad. Sci.* 890, 515–525.
- Lamb, R.J., Preston, K.L., Schindler, C.W., Meisch, R.A., Davis, F., Katz, J.L., Goldberg, S.R., 1991. The reinforcing and subjective effects of morphine in post-addicts: a dose response study. *J. Pharmacol. Exp. Ther.* 259, 1165–1173.
- Lasagna, L., von Felsinger, J.M., Beecher, H.K., 1955. Drug-induced mood changes in man. 1. Observations on healthy subjects, chronically ill patients, and ‘postaddicts’. *J. Am. Med. Assoc.* 157, 1006–1020.
- Martin, W.R., Sloan, J.W., Sapiro, J.D., Jasinski, D.R., 1971. Physiologic, subjective, and behavioral effects of amphetamine, methamphetamine, ephedrine, phenmetrazine and methylphenidate in man. *Clin. Pharmacol. Ther.* 12, 245–258.
- McNair, D.M., Lorr, M., Droppleman, L.F., 1992. Edits manual for the profile of mood states. (Ed.), Edits/Educational and Industrial Testing Service, San Diego, CA.

- Mintzer, M.Z., Griffiths, R.R., 1998. Flunitrazepam and triazolam: a comparison of behavioral effects and abuse liability. *Drug Alcohol Depend.* 53, 49–66.
- Mumford, G.K., Evans, S.M., Fleishaker, J.C., Griffiths, R.R., 1995a. Alprazolam absorption kinetics affects abuse liability. *Clin. Pharmacol. Ther.* 57, 356–365.
- Mumford, G.K., Rush, C.R., Griffiths, R.R., 1995b. Abecarnil and alprazolam in humans: behavioral, subjective and reinforcing effects. *J. Pharmacol. Exp. Ther.* 272, 570–580.
- Mumford, G.K., Rush, C.R., Griffiths, R.R., 1995c. Alprazolam and DN-2327 (Pazinaclone) in humans: psychomotor, memory, subjective, and reinforcing effects. *Exp. Clin. Psychopharmacol.* 3, 39–48.
- Preston, K.L., Bigelow, G.E., 1994. Drug discrimination assessment of agonist–antagonist opioids in humans: a three-choice saline–hydromorphone–butorphanol procedure. *J. Pharmacol. Exp. Ther.* 271, 48–60.
- Preston, K.L., Bigelow, G.E., Liebson, I.A., 1987. Comparative evaluation of morphine, pentazocine and ciramadol in postaddicts. *J. Pharmacol. Exp. Ther.* 240, 900–910.
- Preston, K.L., Bigelow, G.E., Bickel, W.K., Liebson, I.A., 1989. Drug discrimination in human postaddicts: agonist–antagonist opioids. *J. Pharmacol. Exp. Ther.* 250, 184–196.
- 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.
- Roache, J.D., Griffiths, R.R., 1985. Comparison of triazolam and pentobarbital: performance impairment, subjective effects and abuse liability. *J. Pharmacol. Exp. Ther.* 234, 120–133.
- Roache, J.D., Griffiths, R.R., 1987. Lorazepam and meprobamate dose effects in humans: behavioral effects and abuse liability. *J. Pharmacol. Exp. Ther.* 243, 978–988.
- Roache, J.D., Griffiths, R.R., 1989. Abuse liability of anxiolytics and sedative/hypnotics: methods assessing the likelihood of abuse. In: Fischman, M.W., Mello, N.K. (Eds.), *Testing for Abuse Liability of Drugs*, National Institute on Drug Abuse Research Monograph No. 92. National Institute on Drug Abuse, Rockville, MD, pp. 123–146.
- Roache, J.R., Johnson, B.A., Lamb, R.J., Thornton, J.E., Murff, W., Ait-Daoud, N., 2002. A human laboratory model to study internal and external stimulus cues in cocaine reinforcement and relapse prevention. *Drug Alcohol Depend.* 66 (Suppl. 1), 147.
- Rush, C.R., Frey, J.M., Griffiths, R.R., 1999. Zaleplon and triazolam in humans: acute behavioral effects and abuse potential. *Psychopharmacology* 145, 39–51.
- Rush, C.R., Kelly, T.H., Hays, L.R., Baker, R.W., Wooten, A.F., 2002. Acute behavioral and physiological effects of modafinil in drug abusers. *Behav. Pharmacol.* 13, 105–115.
- Sannerud, C.A., Ator, N.A., Griffiths, R.R., 1992. Behavioral pharmacology of abecarnil in baboons: self injection, drug discrimination and physical dependence. *Behav. Pharmacol.* 3, 507–516.
- Schuh, K.J., Griffiths, R.R., 1997. Caffeine reinforcement: the role of withdrawal. *Psychopharmacology* 130, 320–326.
- Schuh, K.J., Schuh, L.M., Henningfield, J.E., Stitzer, M.L., 1997. Nicotine nasal spray and vapor inhaler: abuse liability assessment. *Psychopharmacology* 130, 352–361.
- Schuh, L.M., Schuster, C.R., Hopper, J.A., Mendel, C.M., 2000. Abuse liability assessment of sibutramine, a novel weight control agent. *Psychopharmacology* 147, 339–346.
- Smith, B.J., Jones, H.E., Griffiths, R.R., 2001. Physiological, subjective and reinforcing effects of oral and intravenous cocaine in humans. *Psychopharmacology* 156, 435–444.
- Troisi, J.R., II, Critchfield, T.S., Griffiths, R.R., 1993. Buspirone and lorazepam abuse liability in humans: behavioral effects, subjective effects and choice. *Behav. Pharmacol.* 4, 217–230.
- Walsh, S.L., Preston, K.L., Bigelow, G.E., Stitzer, M.L., 1995. Acute administration of buprenorphine in humans: partial agonist and blockade effects. *J. Pharmacol. Exp. Ther.* 274, 361–362.
- Walsh, S.L., Sullivan, J.T., Preston, K.L., Garner, J.E., Bigelow, G.E., 1996. Effects of naltrexone on response to intravenous cocaine, hydromorphone and their combination. *J. Pharmacol. Exp. Ther.* 279, 524–538.
- West, R., Hajek, P., Foulds, J., Nilsson, F., May, S., Meadows, A., 2000. A comparison of the abuse liability and dependence potential of nicotine patch, gum, spray and inhaler. *Psychopharmacology* 149, 198–202.
- Zawertailo, L.A., Busto, U., Kaplan, H.L., Sellers, E.M., 1995. Comparative abuse liability of sertraline, alprazolam, and dextroamphetamine in humans. *J. Clin. Psychopharmacol.* 15, 117–124.