How Good Are Current Approaches to Nonclinical Evaluation of Abuse and Dependence?

Paul Moser, Toni Wolinsky, Mark Duxon, and Roger D. Porsolt

Porso Int Pharmacology, Boulogne-Billancourt, France

Received July 28, 2010; accepted November 10, 2010

ABSTRACT
Nonclinical assessment of drug abuse and dependence is the subject of several recent regulatory guidelines, which are generally aligned on the methods to be employed. The most direct approach to assessing reinforcing properties of a drug is the self-administration procedure whereby animals can initiate intravenous injections of the test substance, something they readily do with prototypic drugs of abuse. Complications arise because there is no standardized procedure for evaluating substances with differing potencies, reinforcement properties, or pharmacokinetics. Moreover, the choice of training substance, species, and procedural parameters can radically affect the outcome. Apart from the lower cost of rats, primates present several advantages for self-administration studies with similarity to human pharmacokinetics in particular. The most powerful method for assessing similarities between a test substance and a prototypic drug of abuse is the drug discrimination procedure. In contrast to self-administration, drug discrimination is pharmacologically very specific, often reflecting functional activity at receptor level. Dependence is assessed by the occurrence of withdrawal effects on drug discontinuation. Although conceptually simple, interpretation can be complicated by factors such as duration and frequency of administration and observations as well as the choice of end points. Telemetry allows continuous observation of multiple parameters during withdrawal, thereby increasing sensitivity. Presently available tools identify all substances known to cause abuse or dependence, with little risk of false-positives. It remains unclear, however, how predictive these models are with entirely novel substances. Nonetheless, drug abuse/dependence is an area of safety pharmacology where the predictive value of animal models is remarkably high.

Introduction

The nonclinical evaluation of abuse and dependence liability has been the subject of several recent regulatory documents (for details see Moser et al., 2010). The most pertinent of these are the guidelines finalized by the EMA in March 2006 (European Medicines Agency, 2006) and the Draft Guidance issued by the Food and Drug Administration (FDA) in January 2010 (U.S. Food and Drug Administration, 2010). The trigger for this regulatory attention has been the recognition that many commonly prescribed agents can cause unwanted effects on cessation of treatment (Table 1). At the same time, there are increasing concerns regarding prescription drug abuse (Table 2).

Abuse refers to the taking of drugs for nonmedical purposes, usually because of a drug’s positive subjective effects. Dependence refers to the need to continue taking a drug to avoid withdrawal effects on drug discontinuation and is not necessarily associated with any reinforcing properties of the drug. The difference between abuse and dependence can be demonstrated pharmacologically. For example, substances such as the selective serotonin reuptake inhibitors (SSRI) cause clear but mild withdrawal effects but do not possess positive reinforcing properties (Zajecka et al., 1997). Other substances, for example, heroin and cocaine, produce both abuse and dependence. The EMA documents are more concerned with dependence and withdrawal, whereas the FDA documents place more emphasis on abuse. The distinction between abuse and dependence is nonetheless important, because different techniques are required for their assessment.

ABBREVIATIONS: EMA, European Medicines Agency; FDA, U.S. Food and Drug Administration; SSRI, selective serotonin reuptake inhibitor(s); CNS, central nervous system; CPP, conditioned place preference; FR, fixed ratio; PK, pharmacokinetic; GLP, Good Laboratory Practice.
Both the EMA and the FDA recommend a two-tiered approach to nonclinical drug abuse/dependence evaluation. The first tier simply compares the new substance with prototypic drugs. At this stage, common sense is applied to what is known about the test substance. Does it enter the CNS? Does it interact with receptors or uptake sites associated with prototypic drugs? Does it have a therapeutic activity that has been associated with abuse/dependence? If these first comparisons do not suggest possible abuse/dependence liability, the investigations can stop there. If, on the other hand, there are indications of a potential abuse/dependence problem, both agencies recommend a second tier of evaluations where more specific behavioral techniques are applied. Although it is important to consider each substance on its own merits, a possible flow diagram on deciding which tests need to be used for assessing abuse liability is suggested in Fig. 1. The aim of the present article is to critically review the issues and procedures implicated in second-tier evaluations in the light of regulatory recommendations.

**Drug Abuse**

The essence of first-tier evaluations is to compare all that is known of the novel substance with existing abused substances. At this stage, data from radioligand binding studies and simple observational batteries, such as the Irwin procedure (Roux et al., 2004), that evaluate the effects of a substance on behavioral and physiological functions can be usefully employed for detecting similarities between novel substances and prototypic drugs of abuse.

Once a risk of abuse liability has been suggested from the first-tier evaluations, there are a limited number of second-tier procedures that should be considered to evaluate this risk. These procedures cover two aspects: first, whether the test substance possesses positive reinforcing properties, and second, how far the test substance resembles prototypic drugs of abuse. The flow diagram in Fig. 1 outlines a typical decision-making process for determining which tests should be carried out. If the first-tier evaluation suggests possible abuse liability or is not able to rule it out because of some novel activity for which no data are available, the two most important questions are 1) does the substance enter the CNS, and 2) how much does it resemble a known drug of abuse (e.g., because it has a similar mode of action) would probably be evaluated first with drug discrimination and only later in a self-administration procedure.

**Assessment of Positive Reinforcing Properties**

The guidelines mention two major approaches to the assessment of positive reinforcing properties of drugs: conditioned place preference (CPP) and self-administration. These are discussed in more detail below.
Conditioned Place Preference. In terms of procedural simplicity and rapidity, the first approach to assessing positive reinforcing properties is the CPP procedure in which rodents repeatedly exposed to a distinct environment in the presence of a positively reinforcing substance will show preference for that environment when later given a choice (Tschentke, 1998). Under appropriate conditions, CPP can be sensitive to a wide range of substances. Strong reinforcers, such as morphine and cocaine, work reliably and robustly, whereas weaker reinforcers such as benzodiazepines need more specific procedures. CPP thus remains useful for eliminating substances with marked abuse potential during early stages of testing. There are issues with this method for substances such as nicotine and Δ9-tetrahydrocannabinol, however, which tend to have aversive effects on first exposure in this test (Shoaib et al., 1994; Maldonado and Rodríguez de Fonseca, 2002). Priming injections before initiation of place-preference conditioning may allow animals to overcome the aversive effects of initial exposure. Unfortunately, these modifications cannot be applied readily to novel substances, whose initial subjective properties may not be known. Therefore, CPP appears to be a deceptively simple test but cannot be applied blindly to all novel substances.

Self-Administration. The most direct procedure to evaluate the reinforcing properties of a substance is to test whether animals will work (in general, this means to lever press) to obtain the substance. Self-administration assesses the intrinsic rewarding properties of a substance and is not mechanism-based. However, many factors are important in evaluating self-administration of novel substances (for details see Ator and Griffiths, 2003). The following sections concentrate on those areas of particular concern to regulatory authorities.

Choice of training substance. The EMA and FDA guidelines suggest that the training substance should be “appropriate” and that the choice needs to be justified but do not provide specific details. Although cocaine is typically employed, other substances that produce significant levels of self-administration could also be used. Indeed, the available literature data suggest that the choice of training substance may not be the most important variable in a self-administration study but rather that other factors, such as session length or the response requirement for drug infusion, may be more important determinants. This is because, given sufficient time, animals trained on cocaine will self-administer many other substances such as opioids (Griffiths et al., 1981), benzodiazepines (Grant and Johanson, 1987), ketamine, and nicotine (P. Moser, V. Castagné, and C. Froger-Colleaux, unpublished data). The conditions of cocaine training that maximize such transfer have not been systematically studied but are likely to include the dose of cocaine used for training, the fixed ratio (FR) value (i.e., in an operant procedure, the number of responses required to produce an infusion of cocaine), and the length of training sessions. The advantage of using a training drug more closely related to the test substance is that positive effects are likely to be seen over fewer test sessions. A novel test substance will typically be tested for only a limited number of sessions (e.g., 3–5), potentially leading to a false-negative. This could have important consequences if the training and test substances have different properties, such as stimulant versus sedative effects, different pharmacokinetics, etc. (Young and Woods, 1981; Beardley et al., 1990). Unfortunately, choosing an appropriate training drug for a novel test substance is not always straightforward, with the consequence that a greater number of test sessions may be needed to properly evaluate the novel substance. It is possible to perform a self-administration study without training the animals on a prototypic drug of abuse, but in our view, this approach is less useful in the context of safety evaluation. There are several reasons for this. First, it would be very difficult to prove the absence of self-administration as the appropriate test conditions may not have been used; in contrast, with substitution procedures, there is at least a baseline drug demonstrating some level of sensitivity to detect a known drug of abuse. Second, although reinforcing drugs should work in theory under both procedures, there are simply far fewer drugs tested for acquisition as opposed to substitution. In the event of a negative result, the use of an appropriate training substance establishes a level of sensitivity against which to judge the value of that negative result: without such a control, it is more difficult to put a negative result into context.

Procedural considerations. The reinforcement schedule employed, the route of administration, the test session duration, number of sessions, and the doses to be tested are all critical to the outcome of a self-administration study. Each is...
dealt with briefly below (for further discussion see Moser et al., 2010).

**Reinforcement schedule.** Most operant self-administration procedures employ a FR schedule of reinforcement whereby the animal must perform a fixed number of responses to receive an intravenous drug infusion. The number of responses an animal will emit to obtain an infusion generally increases for substances possessing markedly positive reinforcing properties. This is most clearly seen in progressive ratio studies where the number of responses required to earn an infusion is increased until the animal no longer responds. For example, we have evaluated a test substance that clearly supported self-administration in the monkey at FR10 but did not at FR100, whereas cocaine responding persisted beyond FR1000. Thus, the choice of high FR values can decrease the sensitivity of the test.

**Test session duration.** A more subtle parameter is the length of the test session. We have observed that short test sessions in the rat (20 min) limit the ability to detect the self-administration of cocaine compared with saline, whereas it is clearly evident when the test session is prolonged to 2 h. For many substances, animals learn to space out their responses to optimally maintain the positive drug effects. Therefore, a longer test session is required to allow animals to take markedly more infusions of the training substance than those of saline. Several factors, including the pharmacokinetics of the training substance and how much the animals “like” it, probably influence how these infusions are spread out. Test session duration is therefore a particularly important consideration for very long-acting substances, because these may not need to be taken more than once or twice over 24 h to maintain high plasma concentrations. Fortunately, data with the long-acting cocaine analog 2-β-propanoyl-3-β-(4-tolyl)-tropane indicate that this may not be as much of a concern as it at first appears. Although this compound can still produce cocaine-like responding in a drug discrimination assay 24 h after administration (Nader et al., 1997), self-administration in both rats (Roberts et al., 1999) and primates (Lile et al., 2002) is maintained without the need for excessively long test sessions (i.e., substantially less than 24 h, as 6-h long sessions were used in rat). It should be noted that this effect was only clearly seen in primates using an FR schedule, whereas no effect was seen with a fixed interval schedule (Nader et al., 1997).

**Number of sessions.** This parameter has already been mentioned above when discussing choice of training substance, and most investigators generally use five sessions of test substance availability for evaluating its ability to support self-administration. However, data from Mierzewski et al. (2003) suggest that many more sessions may be required to demonstrate a clear increase in infusion number compared with saline sessions. Therefore, investigators must remain sensitive to any indicators showing that a substance may be supporting self-administration and be prepared to extend the number of sessions accordingly.

**Dose-range tested.** Given that abuse liability is unrelated to therapeutic use, the clinical exposure range may not be an important consideration. Therefore, the dose range evaluated in animals needs to extend from a therapeutically “relevant” exposure to the maximal dose at which an animal is still capable of operand responding. Analogous operand procedures, such as lever pressing for food reward or shock avoidance, are the most commonly used techniques for determining the dose at which a test substance starts to have deleterious motor effects that interfere with responding. If the range between therapeutic exposure and such adverse effects is very wide, a larger number of doses should be tested to ensure that the whole range is adequately evaluated. Moreover, dose-response relationships in self-administration studies typically have an inverted U shape, primarily because animals tend to titrate their intake to achieve optimal plasma levels. Therefore, it is also important to test a sufficient number of doses to fully understand this relationship. In certain cases where a long duration of action is expected, the reinforcement schedule should include drug time-out periods to prevent accidental overdosing.

In conclusion, although self-administration is often referred to as the “gold standard” for assessing abuse liability, the procedure itself is far from standard, and different conditions may be optimal for particular prototypic drugs. With a novel substance, it is more difficult to have confidence in a negative result, as the substance may not have been tested under the appropriate conditions. Notwithstanding these considerations, available data suggest that a basic procedure using an FR5 or -10 schedule in cocaine-trained animals has broad sensitivity (see Moser et al., 2010 for more details) and may be an appropriate general training substance in the absence of more obviously appropriate alternatives.

**Assessment of Similarity to Prototypic Drugs**

Although first-tier evaluations can readily suggest similarities between a novel substance and known drugs of abuse on the basis of biochemical or pharmacological properties, the most powerful, selective, and direct procedure is undoubtedly drug discrimination. Animals, by pressing on one of two levers in an operant testing chamber, can show that they are able to discriminate the effects of a particular drug of abuse (training drug) from those of vehicle. It is important to note that drug discrimination is not a measure of abuse liability per se but instead depends on the ability of an animal to identify the subjective interoceptive effects associated with the training drug; these can be positively or negatively rewarding or even neutral (Colpaert, 1987). Humans abuse substances largely for their positive subjective effects; when an abused drug is used for training in animals, there is an assumption that these same subjective effects are involved in its discriminative stimulus properties.

In contrast to self-administration, drug discrimination is pharmacologically very specific. It is therefore a good model to evaluate potential abuse liability associated with particular mechanisms of action. The corollary is that the technique has limited application to substances with novel modes of action. Therefore, administration of potentially abusable substances acting through other mechanisms will result in predominantly saline-lever responding. Drug discrimination can also be used to demonstrate that, despite a high affinity for a particular receptor associated with abuse liability, there is no functional CNS activity in vivo at that receptor.

The critical feature of drug discrimination is the choice of training drug. The more selective the training drug for a particular receptor, the more specific will be the test. An alternative approach is to use the test substance as the training drug, because this would allow numerous comparisons between the test substance and prototypic drugs of

---

**Table:**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reinforcement schedule</strong></td>
<td>Most operant self-administration procedures employ a FR schedule of reinforcement whereby the animal must perform a fixed number of responses to receive an intravenous drug infusion. The number of responses an animal will emit to obtain an infusion generally increases for substances possessing markedly positive reinforcing properties. This is most clearly seen in progressive ratio studies where the number of responses required to earn an infusion is increased until the animal no longer responds. For example, we have evaluated a test substance that clearly supported self-administration in the monkey at FR10 but did not at FR100, whereas cocaine responding persisted beyond FR1000. Thus, the choice of high FR values can decrease the sensitivity of the test. <strong>Test session duration.</strong> A more subtle parameter is the length of the test session. We have observed that short test sessions in the rat (20 min) limit the ability to detect the self-administration of cocaine compared with saline, whereas it is clearly evident when the test session is prolonged to 2 h. For many substances, animals learn to space out their responses to optimally maintain the positive drug effects. Therefore, a longer test session is required to allow animals to take markedly more infusions of the training substance than those of saline. Several factors, including the pharmacokinetics of the training substance and how much the animals “like” it, probably influence how these infusions are spread out. Test session duration is therefore a particularly important consideration for very long-acting substances, because these may not need to be taken more than once or twice over 24 h to maintain high plasma concentrations. Fortunately, data with the long-acting cocaine analog 2-β-propanoyl-3-β-(4-tolyl)-tropane indicate that this may not be as much of a concern as it at first appears. Although this compound can still produce cocaine-like responding in a drug discrimination assay 24 h after administration (Nader et al., 1997), self-administration in both rats (Roberts et al., 1999) and primates (Lile et al., 2002) is maintained without the need for excessively long test sessions (i.e., substantially less than 24 h, as 6-h long sessions were used in rat). It should be noted that this effect was only clearly seen in primates using an FR schedule, whereas no effect was seen with a fixed interval schedule (Nader et al., 1997). <strong>Number of sessions.</strong> This parameter has already been mentioned above when discussing choice of training substance, and most investigators generally use five sessions of test substance availability for evaluating its ability to support self-administration. However, data from Mierzewski et al. (2003) suggest that many more sessions may be required to demonstrate a clear increase in infusion number compared with saline sessions. Therefore, investigators must remain sensitive to any indicators showing that a substance may be supporting self-administration and be prepared to extend the number of sessions accordingly. <strong>Dose-range tested.</strong> Given that abuse liability is unrelated to therapeutic use, the clinical exposure range may not be an important consideration. Therefore, the dose range evaluated in animals needs to extend from a therapeutically “relevant” exposure to the maximal dose at which an animal is still capable of operand responding. Analogous operand procedures, such as lever pressing for food reward or shock avoidance, are the most commonly used techniques for determining the dose at which a test substance starts to have deleterious motor effects that interfere with responding. If the range between therapeutic exposure and such adverse effects is very wide, a larger number of doses should be tested to ensure that the whole range is adequately evaluated. Moreover, dose-response relationships in self-administration studies typically have an inverted U shape, primarily because animals tend to titrate their intake to achieve optimal plasma levels. Therefore, it is also important to test a sufficient number of doses to fully understand this relationship. In certain cases where a long duration of action is expected, the reinforcement schedule should include drug time-out periods to prevent accidental overdosing. <strong>Assessment of Similarity to Prototypic Drugs</strong> Although first-tier evaluations can readily suggest similarities between a novel substance and known drugs of abuse on the basis of biochemical or pharmacological properties, the most powerful, selective, and direct procedure is undoubtedly drug discrimination. Animals, by pressing on one of two levers in an operant testing chamber, can show that they are able to discriminate the effects of a particular drug of abuse (training drug) from those of vehicle. It is important to note that drug discrimination is not a measure of abuse liability per se but instead depends on the ability of an animal to identify the subjective interoceptive effects associated with the training drug; these can be positively or negatively rewarding or even neutral (Colpaert, 1987). Humans abuse substances largely for their positive subjective effects; when an abused drug is used for training in animals, there is an assumption that these same subjective effects are involved in its discriminative stimulus properties. In contrast to self-administration, drug discrimination is pharmacologically very specific. It is therefore a good model to evaluate potential abuse liability associated with particular mechanisms of action. The corollary is that the technique has limited application to substances with novel modes of action. Therefore, administration of potentially abusable substances acting through other mechanisms will result in predominantly saline-lever responding. Drug discrimination can also be used to demonstrate that, despite a high affinity for a particular receptor associated with abuse liability, there is no functional CNS activity in vivo at that receptor. The critical feature of drug discrimination is the choice of training drug. The more selective the training drug for a particular receptor, the more specific will be the test. An alternative approach is to use the test substance as the training drug, because this would allow numerous comparisons between the test substance and prototypic drugs of...</td>
</tr>
</tbody>
</table>
abuse without the need for separate groups of rats trained on each prototypic drug. There may be circumstances where this is a useful approach, for example, if the binding profile indicates that several comparators need to be tested against the test substance.

Drug Dependence

In contrast to abuse that is characterized by continued drug taking, dependence is characterized by the occurrence of withdrawal symptoms on drug discontinuation. Several features of drugs and drug taking are associated with dependence and include frequency and duration of treatment, pharmacokinetic parameters, and the development of tolerance with repeated treatment requiring dose escalation to maintain the pharmacological effect. Drug abuse itself can lead to drug dependence, and continued drug use to avoid withdrawal effects can constitute a contributory factor (West and Gossop, 1994). On the other hand, dependence is not necessarily correlated with abuse liability. Several drugs, for example, the SSRIs, can cause dependence but do not possess positive reinforcing properties (Zajecka et al., 1997).

The general approach to evaluation of withdrawal is to treat subjects for a period of 2 to 3 weeks and then evaluate the occurrence of withdrawal signs for a sufficiently long time after drug discontinuation (at least 1 week). The dosing frequency and route of administration need to be chosen with regard to the pharmacokinetics (PK) of the test substance, and the dose range should be chosen to cover the upper range of therapeutic exposure levels up to maximally tolerated doses.

The pharmacokinetic characteristics of the test substance are important elements in designing a suitable withdrawal study, with regard to dosing frequency, duration of treatment, and the length of the withdrawal observation phase. Withdrawal effects are of particular concern for short-acting substances. The symptoms tend to be more severe than for substances possessing a long half-life where the exposure tails off much more gradually, thereby attenuating withdrawal effects (Farré and Camí, 1991). Many substances under development are being designed with long half-lives to allow once-a-day administration. Under these circumstances, withdrawal effects are likely to be more subtle and require an experimental approach that is more sensitive to small changes.

Measures of Withdrawal

The withdrawal effects of opiates, amphetamines, and barbiturates have been extensively characterized (West and Gossop, 1994). Simple behavioral observation can identify many of these signs that include tremor, wet dog shakes, and hypothermia as well as reduced Food intake and body weight gain.

The regulatory authorities are becoming increasingly concerned with more subtle withdrawal effects, which may be a feature of many CNS-acting substances as listed in Table 1. For example, the reported effects of withdrawal from SSRIs, in contrast to opiates, are mild and include dizziness, headaches, anxiety, and fatigue (Zajecka et al., 1997). The EMA document calls for new approaches to make testing more sensitive to such effects, for example, by using tests for anxiety and cognition. Unfortunately, these procedures are often time-consuming and difficult to carry out repeatedly because the behavior of animals during subsequent exposures is different from the first (Lamprea et al., 2000).

A novel approach is the use of telemetry, which can continuously monitor several parameters (body temperature, locomotor activity, heart rate, and blood pressure, respiratory rate) in conjunction with the more classic withdrawal measures of food intake and changes in body weight (Chan et al., 1999; McNally and Carrive, 2006; Moser, 2009). A further advantage of telemetry is that it permits monitoring 24 h per day, thereby decreasing the possibility that important events will go undetected.

General Considerations

The above discussion has reviewed the different procedures recommended by the regulatory authorities for the nonclinical evaluation of abuse and dependence liability. To conclude our treatment, more general topics covering the whole field are discussed in the sections below.

When Should Abuse/Dependence Studies Be Carried Out?

Both the EMA and the FDA recommend that abuse/dependence studies be performed during phase II clinical studies. Furthermore, it is recommended that PK data from phase II trials be available so that the exposure levels used are relevant to the clinical situation. The main concern here is to prevent patients from being exposed to a potentially abuse-or dependence-inducing substance during large-scale phase III trials. On the other hand, from the pharmaceutical industry perspective, there are good reasons for carrying out such tests at an earlier stage to avoid the time and expense of clinical development. For example, the value of a project/therapeutic target might be greatly reduced if the substance were to be scheduled.

Choice of Species?

The rat is a much less expensive experimental subject than the primate. This would present an argument in favor of the rat. Are there other arguments favoring the use of one species over the other?

With regard to self-administration, the EMA guideline document (European Medicines Agency, 2006) clearly recommends the rat. In contrast, the more recent FDA draft guidance (U.S. Food and Drug Administration, 2010) does not favor one species over the other. The EMA’s preference for the rat is no doubt partly driven by ethical concerns aimed at reducing the use of primates in drug evaluation. In addition, there exist extensive data indicating no major species differences in the predictive value of self-administration procedures; both are highly correlated with human abuse liability (Brady and Fischman, 1985).

Whatever the correlation between self-administration in rats and primates, other factors militate in favor of the primate (see Weerts et al., 2007 for an extensive review). Perhaps the most important is the greater similarity of pharmacokinetics between primates and humans compared with rats (Ward and Smith, 2004a,b). As mentioned above in the discussion of session duration, the pharmacokinetics of the test substance can play a major role in determining how readily it will be self-administered. Likewise, the affinity or selectivity of the test substance for different target sites, as well as the
distribution of those target sites, is likely to be closer to humans in the primate than in the rat. These two elements, selectivity and pharmacokinetic profile, should therefore be the main considerations in choosing the most appropriate species for testing. Such data will usually be available when abuse liability evaluation is being considered.

Dose-limiting side effects should also be considered when assessing abuse potential because these can play a role (e.g., as with bupropion, see Table 3). As with pharmacokinetics, drug-induced symptoms in the primate better model what will be seen in humans. For example, sedation, drowsiness, excitation, aggressiveness, or motor incoordination, all of which may be dose-limiting for abuse, are more readily translatable to humans from primates than from the rat and a meta-analysis of toxicity studies has found that primates are better predictors of human adverse events than are rodents (Olson et al., 2000).

A final factor favoring the primate over the rat is the manner in which self-administration experiments are conducted in the two species. Primates are typically used on a repeated basis, frequently over several years to assess the abuse liability of a wide range of different substances. In contrast, the rat is used over much shorter periods (several months), usually for the evaluation of a single substance. The reason is partly technical because of the difficulty of maintaining the patency of rat catheters over long periods and partly physiological; the rat continuously gains body weight over time thereby vitiating within-subject comparisons. Within-subject comparisons are therefore more readily realizable in the primate. Moreover, the fact that the primate can have experience of a wide range of different substances brings it closer to the human situation where the street user typically experiments with numerous drugs.

With regard to drug discrimination, there does not seem to be any clear advantage in using primates. Active doses and PK considerations are not critical for obtaining stable discriminative control, and the behavioral symptom profiles are less relevant because drug discrimination experiments are carried out at doses below those having observable behavioral effects. Again there seems to be a high correlation between results obtained in the two species (Gerak and France, 1996). Therefore, with drug discrimination procedures, cost considerations would argue in favor of the rat.

For drug dependence studies, cost considerations would probably also predominate. In contrast to self-administration where primates are used repeatedly, drug dependence studies usually require the use of drug naive and, therefore, highly expensive animals. It is nonetheless true, as indicated above, that dosing and PK considerations are critical for the study of withdrawal phenomena, and primates thus are more likely to show greater similarity to man.

A determining factor in the choice of species is the time at which the studies are undertaken. If the drug abuse/dependence assessments take place during the early stages of a drug discovery program, cost considerations would favor the rat. At later stages of drug development, for example, during clinical testing, the arguments for primates indicated above would come more into play.

**GLP or Not GLP?**

At the time of writing, there was no strict requirement from the regulatory bodies to carry out abuse liability studies.

### Table 3

**Some possible discrepancies between animal and human data**

<table>
<thead>
<tr>
<th>Compound</th>
<th>Human Data</th>
<th>Animal Data</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modafinil</td>
<td>Identified as reinforcing by drug users but limited evidence for systematic abuse</td>
<td>Modafinil is positive in numerous animal models, including self-administration, place preference, and discrimination</td>
<td>Evidence for abuse in humans is weak but possibly because other stimulants are available more readily and more cheaply</td>
</tr>
<tr>
<td>Bupropion</td>
<td>Only a few case histories of abuse</td>
<td>Active in numerous models of abuse, including self-administration</td>
<td>Bupropion is convulsive at lower doses than lead to positive reinforcing effects</td>
</tr>
<tr>
<td>Salvinorin A</td>
<td>Widely abused</td>
<td>Very little evidence supporting abuse. Shows opiate ß-agonist activity in discrimination tests in rats</td>
<td>Product cheap and freely available so even if effects are very mild there are few barriers to abuse</td>
</tr>
<tr>
<td>Performance-enhancing agents (e.g., steroids, EPO, etc.)</td>
<td>Widespread abuse in particular areas (e.g., sports)</td>
<td>Very limited data available under specific conditions but suggest that steroids might be weakly reinforcing (e.g., Wood, 2004; Jorge et al., 2005)</td>
<td>Animal models are designed to detect abuse related to positive reinforcing effects. If those positive reinforcing effects are due to lifestyle effects, then they fall outside the scope of testing</td>
</tr>
<tr>
<td>Erectile dysfunction treatments (e.g., sildenafil (Viagra))</td>
<td>Widespread abuse</td>
<td>None reported</td>
<td></td>
</tr>
</tbody>
</table>

EPO, erythropoietin.
to full Good Laboratory Practice (GLP) compliance, although as with all safety pharmacology studies, GLP compliance adds a level of quality control. The requirement for carrying out these studies to “high scientific standards” is suggested to be more important than GLP compliance in the EMA guidelines. For the moment, the ICH S7A document (European Medicines Agency, 2006) places abuse/dependence liability assessment under secondary or follow-up safety studies and requests that they are carried out in compliance with GLP “to the greatest extent possible.” Both the EMA and FDA refer to the ICH S7A text in taking their stance on GLP compliance. In practice, as more and more laboratories are able to offer these studies under GLP conditions, it is likely that GLP compliance will become the norm.

**Conclusions**

The methods for assessing drug abuse liability discussed above can identify virtually all substances known to possess abuse liability in humans, with little risk of false-positives. Apparent discrepancies between human and animal data for some substances can usually be explained by other characteristics of the substance (Table 3) or because, ultimately, abuse liability is the sum of many factors, including the degree of substance reinforcement, its side effects, its availability, and its cost. Animal models can only really assess the first one or two of these. In addition, these models have so far only shown their merits with substances already known for street abuse. How well the methods will perform with completely novel substances remains to be seen, and more work is surely needed to improve sensitivity to some classes of compound, such as cannabinoids and glutamate antagonists. The fact that the tests discussed here are mostly independent of any particular mechanism of action or interaction with particular neuronal systems gives grounds for confidence that they will continue to be predictive. For this reason, they are expected to remain the core of abuse and dependence liability assessment for some time to come. Increasing our understanding of how the variables discussed above can affect our ability to detect various classes of abused substance is certainly needed, as well as more guidance on what level of sensitivity is required. There are also some areas that have not been investigated to any great extent, such as how these models can be applied to juvenile subjects. Abuse of medication is increasingly being seen in younger people, and there are concerns over its association with development of psychiatric disease (e.g., Saban and Flisher, 2010), whereas the vast majority of nonclinical studies are carried out in postpubertal animals (e.g., Smith, 2003). It should also be noted that many substances are abused for reasons other than their positive reinforcing properties, e.g., for their effects on erectile dysfunction, body weight, muscle development, or motor performance. It is unlikely that these will be detected via nonclinical testing of the kind described here. Nonetheless, the primary concern of abuse liability assessment in the context of safety pharmacology is to detect risks unrelated to the medical use of the substance.

Withdrawal and dependence are less well studied, and techniques are evolving all of the time. Nonetheless, we believe that current nonclinical procedures, if designed appropriately with the pharmacology of the test substance in mind, are capable of detecting most effects of clinical concern. It should be noted, however, that there are few nonclinical studies describing clear withdrawal effects in animals with SSRIs, the class of substance that was at the origin of the EMA’s concern. Overall, we believe that available nonclinical methods for drug abuse/dependence assessment constitute an area of safety pharmacology where the predictive value of the animal models is remarkably high.

**Acknowledgments**

We thank Dr. Roger D. Forsolt for comments on the manuscript and Ymame Boujbar for expert secretarial assistance.

**Authorship Contributions**

Wrote or contributed to the writing of the manuscript: Moser, Wolinsky, and Duxon.

**References**

Moser et al. 594

---

**References**


