There can be no doubt that many drugs have adverse effects on driving, i.e. a special form of behavioral toxicity (BRADY, 1959) concerning overt behavior, such as performance, and covert behavior, such as mood and motivation. Psychopharmacological research in road safety is mainly directed towards the following questions:

a) which drugs increase driving-risk,
b) what are their effects,
c) in whom and
d) under what conditions do these effects occur, and
e) how great is the drug-produced increase of risk.

Results in a) - e) should serve as a basis for decisions and countermeasures. While there are numerous studies dealing with a) and b) there are only a few concerning c) and d), and nearly no studies on the extent of increase of actual risk produced by drugs other than alcohol. Behavioral toxic effects have been measured (by means of diagnostic psychological methods) but these results cannot be translated into quantification of risk in real life situations with any degree of accuracy. So far there has been no coordinated approach for testing of drug effects and no risk assessment model has been developed for objective risk prediction (JOSCELYN, 1975).

"The drug-driving risk may basically be defined as the likelihood that a road accident will occur, due to abnormal driver behavior caused by a drug, all other things being equal" (MAICKEL, 1975).
In this context the fact that abnormal driver behavior may produce a situation in which not the impaired driver himself but others are involved in a crash ought to be mentioned. This represents a further complication in evaluating drug-procued increase of risk.

To evaluate drug-induced dysfunction of the system driver - environment - vehicle, systems analyses and task analyses are necessary. HELANDER (1973) presents a theory on some interactions within this system describing the probability of homeostasis or collapse as a function of the difference between the performance quality of the driver and the performance demand of the environment (see also BLUMENTHAL, 1968).

However, satisfactory analyses of the driving task itself has not been achieved so far. We know a number of factors, yet their relative importance and their interactions are unknown to a considerable extent.

2. VALIDITY, RELIABILITY, OBJECTIVITY

In this dilemma of evaluating drug-effects on a task not yet satisfactorily described various approaches have been tried: laboratory studies using a wide range of tests, simulator studies and in-vehicle studies on closed courses or in normal traffic.

Common to all of them are the problems of validity, reliability, objectivity and sensitivity.

WITTENBORN (1971) states, "Any measure which changes in consequence of the medication has validity for that medication". Yet, whether this measure is also valid for real-life behavior under study has to be established over and above this. One of the basic difficulties is related to the question of valid criteria. The use of the number of reported accidents as a criterion to validate behavioral measures is highly questionable because their occurrence is infrequent, they are dependent on the interaction of a great number of factors (environmental, vehicular and driver variables) and accident-reports are highly unreliable in themselves (KLEBELSBERG et al. 1970, EDWARDS et al., 1969). On the other hand HÄKKINEN (1957) shows a high reliability of number of accidents in a homogeneous group over a long period of time. In not homogeneous groups variables of exposure interfere.
Thus a comparison between experimental data and number of accidents might lead to an under-estimation of validity and there is no way to decide whether a low correlation is due to deficiencies in the criterion or to inappropriate behavioral measures. Therefore, standardized techniques for observing and recording driver behavior were developed (e.g. KLEBELSBERG et al., 1970, HÖFNER, 1967, KROJ & PFEIFFER, 1973, QUENAULT & FUHRMANN, 1969) and their results used as a criterion. During the last few years observations of conflict-situations or near-accidents has gained increasing importance. This approach certainly has its merits if tests, simulator studies, or driving on test-courses are validated against it: A sample of real-life behavior can be rated according to a number of variables which then can be compared to patterns of performance. There are, however, inherent problems in this method of validation: In the techniques described by KLEBELSBERG (1970), KROJ & PFEIFFER (1973), and QUENAULT & FUHRMANN (1969) two observers ride with the subject and thus by their presence might influence his performance. The authors maintain that if the period of observation is long enough these effects will disappear. The method used by HÖFNER excludes these difficulties since drivers are not aware of being followed by observers but practical problems, such as waiting for subjects to drive, occur. Also, there is no way to standardize the route on which observations are made and observations are restricted to certain variables. Conflict techniques seem promising, but here, too, is no way to standardize traffic situations. There are test-batteries currently in use which have been validated (using accident data or standardized observation of driver behavior as criteria) although correlation-coefficients hardly ever exceed 0.5 (OSWALD, 1969, KLEBELSBERG et al., 1970). This applies also - to a certain degree - to simulators. (MOSKOWITH, 1975, SCHUBERT, 1965).

JANKE (1961) points out another problem of validity: The factorial structure of a test may be changed by the administration of a drug. This was shown in a study on the influence of LSD on the factor structure of an intelligence test (LIERNERT, 1959). It still remains to be shown whether drugs with less drastic effects, such as tranquillizers, also change the factorial structure (and, therefore,
the factorial validity) of tests and whether this applies to tests other than intelligence-tests (e.g. psychomotor or perception tests). Reliability is another condition that must be fulfilled by behavioral measures of drug effects on driving. In this context practice or learning effects will have to be considered, too. Behavioral measures also must be objective, i.e. they have to be repeatable and lend themselves to easy and unequivocal quantification (LEHMANN, 1960). Validity, reliability, and objectivity cannot be seen independently of each other. It is still an unsolved problem which of these conditions should be given priority, yet it is generally agreed that for an experimental behavioral method acceptable data on all three - validity, reliability and objectivity - should be provided.

3. SENSITIVITY AND PRACTICAL SIGNIFICANCE
Provided that behavioral measures are satisfactorily valid, reliable, and objective there is still the question of their sensitivity and of the statistical vs. practical significance of drug-effects on performance measures. Some authors maintain that techniques for drug studies should be at least sensitive enough to show the effects of small doses of alcohol (0.4 - 0.5 % BAC) yet the effects of other drugs may be qualitatively different from those of alcohol (JOSCELYN, 1975). So far there are no generally applicable or accepted rules for sensitivity of techniques.
This deficiency is closely related to the question of practical significance: Whereas there are definitions of statistical significance, a statistically significant drug effect on experimental behavioral measures does not automatically imply that this effect is of practical significance, that is whether it would interfere with driving to an extent which would substantially increase risk (COLE, 1960, KLEBEL, 1975). We are as yet far from a generally accepted definition of substantially increased risk.
Statistical significance is necessary but not sufficient for establishing practical significance, it gives no information upon the extent of the drug effect. Statistically highly significant results may be obtained either from a large effect found in a small
sample or from a small effect in a large sample. COHEN (1962) suggests a scheme of "large", "medium", and "small" treatment effects, defining a mean difference of one standard deviation unit as "large". This seems questionable in view of the considerable effect of "non-specific factors" on drug-response (RICKELS, 1968, DOWNING & RICKELS, 1970). Moreover, it must be borne in mind that although some factors are significantly impaired other factors may stay intact or even be enhanced so that compensation is possible. In some cases the subject's awareness of impairment of certain functions (e.g. depth perception) may produce more caution (COLE, 1960) or more motivation to overcome the impairment. As yet we know too little about the mechanism of compensation and its limits, e.g. minimum requirements that have to be met or which kinds or patterns of impaired functions can or cannot be compensated for, and for how long compensatory mechanisms can be kept up. Some authors maintain that through reactively increased mental exertion under moderate doses of depressants performance may be enhanced (DÖKER, 1963, FORTH, 1966, JANKE, 1964). BIEHL (1972) could not substantiate this hypothesis. However, recent studies indicate that subjective perception of impairment is important for compensation. This ties in well with the risk-compensation theory developed by WILDE (1976) and KLEBELSBERG (1971).

4. POPULATION AT RISK AND SAMPLING

The adequate selection of subjects is another problem which cannot be overlooked: If results obtained from the experimental sample are to be generalised to a specific population, it is necessary to select subjects in such a way that they truly represent this population. The population in question here is that of all drivers. On the one hand this means that samples of patients who are unfit to drive anyway (e.g. hospitalized chronic schizophrenics) are certainly inappropriate, on the other hand studies conducted using young healthy volunteers (in many cases college-students) are of limited value for the investigation of side-effects of therapeutic drugs.
or are impossible sometimes. This last issue becomes especially clear when an antihypertensive agent is under study. When we bear in mind that morbidity in the group of college-students is low compared to that in the group of over 50 year olds, generalization of results seems questionable: Older people take more therapeutic drugs and may react to them in a different way, while younger people tend more to the use of illicit drugs.

There is also some evidence that psychoactive drugs have different behavioral effects on patients (regardless of age) than on normals and may even improve performance on a driving battery (Smith et al., 1958, OECD, 1977). However, samples of young healthy volunteers certainly have advantages (e.g. no additional medication, no impairment by varying degrees of illness, greater homogeneity, usually good motivation and cooperation) and should not be omitted in the evaluation of a new drug.

Reitan (1960) suggests inducing experimental stress or conflict in such a sample. This method has been used by several authors among them Janke & Glathe (1964) who introduced noise as a stressor, or Frankenhaeuser (1974) who used experimental conflict-situations. Also, drug effects on experienced as well as inexperienced drivers should be studied.

5. DESIGN OF STUDIES

There is general agreement that studies should be conducted double-blind, using either placebo, or a reference drug, or both for controls. Which of these methods is applied depends on ethical considerations (if patients are used) and also on the hypothesis to be tested. However, a placebo-group should be used whenever feasible because of the unspecific drug effects which cannot be controlled otherwise and also because a special sample in a certain setting might react to the reference-drug in an uncharacteristic way. It is necessary to plan the manner of statistical analyses already when experimental design is planned. Since interactions with non-drug variables in the individual as well as in the setting are common (Levine, 1971) experimental design should be arranged in such a manner that these interactions can be controlled or accounted for. Many tests show marked practice-effects which should be taken into
consideration, too. Also, the fact that drug-response is not linear as a rule (BIEHL, 1972) and many behavioral measures do not follow a Gaussian distribution (e.g. reaction times, errors in concentration tests) should be taken into account. UHR (1960) and LEVINE (1971) provide suggestions on design and statistical analyses. FORTH (1966) describes a model of the relationships between different doses of depressants and change in performance. He maintains that moderate doses of such drugs may improve performance while performance is progressively impaired by medium and higher doses (see also DI MASCIO, 1964).

MILNER (1975) states that most psychoactive drugs have a continuous logarithmic normal curve for their dose response. LIPPERT (1959) describes the following general rules for behavioral response to dosage:

a) Small doses have different effects upon different personalities
b) High doses: delirium as a result of high toxicity
c) The range of typical drug effects lies between a) and b)
d) Every drug has a dosage when it will not show any effects (except unspecific ones).

This means that before the effects of a drug on driving performance is studied dose-finding studies have to be carried out so that "small", "medium" and "high" doses can be defined, emphasizing normal therapeutic doses. Also, a generally accepted model of drug effects on behavior is still to be established.

The effects of a specific dose also depend on the type of task: moderate doses of depressants impair performance only on highly complex tests, whereas higher doses affect simpler functions, too. There is considerable evidence that serum levels after a particular dose vary enormously (e.g. NASH, 1964, MILNER, 1975, SELLERS, 1975, OECD, 1977). SELLERS reports that correlations of serum concentration and effects are better than those established for dose and effects for phenothiazines, nordiazepam, nortriptylin, desipramine, and LSD among others, and suggests that analyses of plasma concentrations should be included in future studies of drug effects on driving skills. This method, however, is very complicated and expensive, and adequate analytical methods are not yet available for all drugs including their active metabolites. Moreover, too
little is known about pharmacokinetics of new drugs especially and for some drugs, at least, maximum effects are not simultaneous with maximum plasma concentrations (JOSCELYN, 1975). Therefore, NASH's suggestion to adjust drug effects for body-weight or surface area employing covariance analysis seems to be a feasible compromise: a substantial reduction of experimental error may be achieved, and at the same time, the practical value of the study may be higher because as a rule - at least for outpatients - recommended therapeutic doses usually do not refer to individual body-weight or surface area, or the resultant plasma concentration. This method takes into account only the volume of body fluids; errors due to other internal factors, such as stomach emptying time or variations in metabolism are unaffected. Special emphasis should be given to therapeutic doses but low and high doses have to be studied, too, because in low doses there may be negligible effects or even enhancement of performance, and regarding high doses it may well be that abusive use of the drug is the problem in some cases (JOSCELYN, 1975). The latter issue would need clarification by epidemiological studies, too. In acute studies time after application is of great importance: According to the pharmacokinetic characteristics of the drug behavioral tests have to be timed in such a way that peak effects will be shown, so that the decline of drug effects during time does not influence performance on various tests of a battery. NASH (1964) suggests some forms of administration which are designed to maintain a certain drug level throughout the tests. But maintaining the concentration of a drug at a uniform level does not in itself insure that the effects of the drug will be maintained at a uniform level because short-term adaptation of the central nervous system may set in (NASH, 1964).

Bias due to temporal variation can be kept minimal by either combining symmetrical placement of equivalent subtests with the presentation of tests in opposite orders (NASH, 1964) or testing different functions on different days (LIPPERT, 1959). When this latter procedure is adopted intervals between test-days must be chosen in such a way that cumulative effects are ruled out. The most common approach is testing two or three hours after drug administration. Yet psychoactive drugs or their active metabolites
usually stay in the system much longer than that, also hangover effects have been observed (WALTERS & LADER, 1971). Therefore, even after application of a single dose, repeated testing up to 3 or 4 days can be advisable (JOSCELYN, 1975). PÖLDINGER (1965) compares the effects of twelve different psychoactive drugs with placebo testing repeatedly during 7 hours. He finds typical time course variations for the effects of several groups of drugs.

The majority of psychoactive drugs are used in chronic application more often than in single doses. Yet the majority of studies using normal subjects deal with acute rather than chronic applications. Physical and behavioral tolerance (HARVEY, 1971) (or possibly the reverse) are likely to develop. Therefore, more studies of the effects of chronic administration on driving skills in the patient population under therapy (SELLERS, 1975) as well as in normals should be conducted. Whenever possible testing should go on for an adequate period after discontinuation of medication to test for after-effects or withdrawal effects.

6. OTHER FACTORS INFLUENCING RESULTS OF BEHAVIORAL DRUG STUDIES

In addition to type of drug, dose, and timing there are a great many factors that may create either intraindividual or interindividual differences in drug response (JANKE, 1965), or both. These factors may either be inherent in the individual itself or consist in external variables. For reasons of economy and simplicity the latter classification will be used here.

6.1. Factors in the individual

Certainly sex and age play an important role and have to be taken into account in the experimental design.

Personality factors have a marked influence on drug effects (JANKE, 1960, 1964, 1965, EYSENCK, 1963) mainly when doses are moderate while differences tend to disappear with increased doses (BIEHL, 1972, KORNETSKY, 1957).

DI MASCIO (1960, 1963, 1968) found differential responses (physiological, behavioral, and subjective) for two personality types and concluded that individuals whose personality was organized about active mastery of the environment (by athletic prowess, hostile
outbursts, or extrapunitive acts) found sedative drug actions ego-threatening and their intellectual functioning was impaired. In contrast intellectual and anxious individuals experienced reduction of anxiety and intellectual performance was enhanced. FROSTAD et al. (1966) obtained similar results using a concentration test.

There are numerous studies using a wide variety of personality factors (e.g. FAHRENBERG & PRYSTAV, 1969, Mc DONALD, 1967, FORREST, 1967, MUNKELT, 1964, JARKE, 1960, 1964, 1965, EYSENCK, 1963) with uniform results; however, they provide experimental evidence that personality characteristics influence drug effects. A valid model of the interactions personality - drug has not been established so far. Inspite of this there is one factor whose influence upon drug effects has been established fairly well: The level of central nervous activation has been found of marked importance both for drug effects and for performance (HARVEY, 1971, SHAGASS, 1960, EYSENCK, 1957, FRANKENHAEUSER, 1975, HELANDER & SÖDERBERG, 1973, HELANDER, 1973, LEGEWIE, 1968).

Optimal performance is reached when cortical arousal is medium, whereas impairment is observed when central nervous activation is either low or high. Similarly, mild sedation will produce better functioning in persons whose level of arousal is too high and impairment in subjects with low activation. Stimulants will have reverse effects. LEGEWIE (1968) points out that this has to be seen also in connexion with the task in question: level of activation can be too high, too low, or optimum for the specific task. Optimum level of activation decreases as complexity of the task increases.

Another important factor are placebo-reactions. While some authors maintain that there are "placebo reactors" who could be screened out to reduce experimental error LASAGNA (1971) doubts that individuals are either "placebo reactors" or "non-reactors". Whether or not reaction to placebo is a fairly constant personality trait has not been satisfactorily explored yet. It is certain, however, that placebo reaction depends to a considerable degree on the situation and the expectations of subjects as has been shown by BEECHER (1968) who found that enthusiasm versus skepticism may exert a powerful influence, or FRANKENHAEUSER et al. (1964) and
KOWAR (1973) who found that the effects of information and subsequent expectations were stronger than the drug effects. Therefore it will be necessary to create equal expectations in experimental groups as well as in control groups. One approach is not to give any information on the substances under study but this might be difficult under the informed consent law and ethically questionable. Another possibility is to specify the type of active drug and give no information that placebos will be used. This might be ethically acceptable in normal subjects, and, moreover, expectations may be kept more uniform by this method than by giving no information at all.

Another advantage of this may be seen in the fact that under normal therapeutic conditions drugs are taken with certain expectations as to their effects. In order to be able to control better for placebo-effects a placebo wash-out period (under the same information and expectations as in the main drug-trial) may be employed and results from this may be taken into account by appropriate statistical methods.

The psychological effects of a drug are also influenced by the subject's pharmacological history (GOLDBERG, 1943, JELLINEK & McFARLAND, 1940, GOODMAN & GILMAN, 1955). It might, therefore, be advisable to use only subjects with a specified amount of experience with the drug in question or similar drugs, or degree of experience could be controlled by statistical treatment of data.

Predrug level of functioning has been shown in a number of studies to influence drug effect. There are not many studies dealing specifically with this issue and results are still controversial:

Whereas SCHNEIDER (1960) found that chlorpromazine prolonged the reaction time most in those subjects who before drug administration were fast reactors, and a stimulant speeded reaction time more in slow reactors, BROEREN & SCHMITT (1965) found enhanced performance in a concentration test in subjects with initially good concentration under an antidepressant and no effects on subjects with poor pre-drug performance. This indicates that initial values may be affected in a different way depending on the drug used and on the function tested. Current psychological and physical state of the subject before drug administration and also during the trial (in long-term studies, especially) may also influence results (HARVEY, 1971). Changes not
due to the drug but e.g. due to meteorological and other environmental factors or to rhythms established within the organism often cannot be controlled. The subject's trend of thoughts and his emotions may also bring about alterations in performance (LEHMANN, 1964). Also, motivation and set may play an important role (JANKE, 1964, 1965) and may vary during the course of a test (FISKE, 1964) and be changed through drug effects (BÄUMLER, 1975).

6.2. External variables
Setting may influence test performance considerably. Group situations may e.g. evoke sociability or individual assertiveness (CARTER, 1954, BORGATTA et al., 1958) and thus influence behavior and obscure drug effects. Also, group settings may influence drug effects, especially facilitating them if all members receive the same drug (cf. effects of alcohol or hashish), or the apparent effects on a subject may be low if others are given different treatments, or placebo-effects in the control group may be increased by subjects of the experimental groups commenting on their experiences with the drug. There is as yet a deficit of experimental studies and no model for evaluation of these effects in drug trials. At any rate, communication between subjects should be minimized by the experimenter (NASH, 1964). The type of tasks used will have marked influence on the results of a behavioral drug study, LEVINE et al. (1975) showed this for alcohol. As stated above, complex functions will be affected by moderate doses (or less powerful drugs) while simple functions will not be changed unless higher doses (or more powerful drugs) are used. Performance on highly practiced tasks will be less likely to be impaired. Numerous studies show typical performance changes characteristically brought about by particular drugs (a review is given by ORZACK, 1975). For instance there will be no point in observation of a few minutes performance on a test-course if vigilance is the function most impaired by a drug. Therefore, either a very wide range of tasks has to be used or there must be sufficient information on characteristic drug effects previous to planning the study.

Practice effects are high in many types of tasks. In order to minimize their influence testing can be conducted near the flat por-
tion of the practice curve or, when feasible, equivalent forms of
tests can be used (NASH, 1964). This will of course not be necessary
when learning is a variable under study.
The person of the experimenter also is an important factor in drug
studies even when he has no specific expectations of drug-effects.
His personality, sex, age may influence test results. Therefore,
whenever possible, there should be no change of experimenters
during a study.
Stress may vary drug response considerably as was shown e.g. by
(1974) showed that stress may also manifest itself on performance
decrements after the termination of the stressor. Also driving
behavior is susceptible to stress as was shown, among others, by
HELANDER (1973), HELANDER & SÖDERBERG (1973) and JOHANNSEN (1975).
Therefore, an inclusion of behavior under stress in studies of
drug effects on driving seems advisable.

7. CONCLUSIONS AND RECOMMENDATIONS
Returning to the questions posed in the beginning of this paper we
can state that by behavioral methods drug effects on driving so far
can be established qualitatively only. A quantitative estimation
of increased risk is not possible yet.
Pre-marketing screening should be done using a driver battery of
laboratory tests as well in acute as in chronic studies on normals
so that a wide range of functions is studied. Then drug-effects
that need further clarification can be pointed out and studied by
a variety of experimental methods, including simulation and
in-vehicle experiments and a profile of drug effects should be
set up. Also, effects on patients ought to be investigated then.
This is indispensable for the practical value of the findings.
The following crucial issues are recommended for priorities for
further research:
- Task analysis of driving behavior and systematized description
  of factors stating their relative importance and interactions.
- Validity and reliability of experimental behavioral methods.
- Standards for sensitivity of behavioral methods.
o Quantitative standards for practical significance of results.
o A model of drug effects on behavior and interactions with factors in the individual and external factors.
o Description of valid sampling procedures for the population under study and the hypotheses to be tested.
Thus a more systematic approach to the experimental investigation of drug effects on driver behavior would be possible.

List of references available from the author upon request.