ELSEVIER

Strategies to optimize the validity of disease models in the drug discovery process

Frank Sams-Dodd

Bionomics Europe, Les Algorithmes, rue Jean Sapidus, Parc d'Innovation, 67400 Illkirch, France

Models of human diseases are necessary for experimental research into the biological basis of disease and for the development of treatments. They have an enormous impact upon the success of biomedical research. However, in spite of this, a consistent system for evaluating, expressing and comparing the clinical validity of disease models is not available. The purpose of this paper is, therefore, to provide a theoretical discussion of the concepts behind disease models and to develop a terminology and a framework to analyze and express the clinical validity of disease models.

Models of human diseases are necessary for experimental research into the biological basis of disease and for the development of treatments. In the pharmaceutical industry, disease models are used for target identification and validation, drug screening and proof-of-principle studies, and they can be considered to be both contributors to, and restrictors of, the drug discovery process. An invalid disease model can lead the industry in a wrong direction, thus wasting time and significant investment. A new model with increased validity over existing models can mean the discovery and clinical testing of entirely new approaches to disease treatment. Disease models are, therefore, crucial to the drug discovery process and have an enormous impact on the success of the industry and advances in healthcare.

Ideally, a disease model should fully reproduce the clinical condition in a system that can be used for research and drug discovery. However, in reality, disease models usually only model certain aspects of clinical symptomatology, and because only rarely is the aetiology of diseases well understood, the induction of the disease state in the model can differ from the clinical condition. A disease model, therefore, has many limitations and it is necessary to appreciate these to use a disease model correctly and to be able to select between disease models for a given drug discovery programme. The present review firstly, based on the human disease, defines the components that must be included in a disease model, then describes the disease model and, finally, presents a system for evaluating and comparing disease models. The system will be applied to screening systems that are used frequently, and the special case of the proof-of-principle study will be discussed.

Corresponding author: Sams-Dodd, F. (fsams-dodd@bionomics.com.au)

The human disease

The human disease state can be described by five basic components: the disease, the patient, the treatment, the symptomatology and the diagnostic tool (Figure 1a).

Disease

Disease can be caused by a multitude of factors. The cause can be genetic factors, such as mutations – for example, a familial mutation causing Alzheimer's disease (AD) – or disease risk genes that confer an increased susceptibility to develop a particular disease under certain conditions. Disease can also be caused by extrinsic factors, such as infectious agents and parasites, toxic agents, stress or traumatic events (e.g. a concussion), or intrinsic factors, such as autoimmune responses and other malfunctions in the normal physiology of the organism. In most cases, a disease will originate from a combination of extrinsic and intrinsic factors that, in special circumstances, lead to the induction of the disease process; for example, studies suggest that schizophrenia can be caused by a viral infection during pregnancy or by hypoxia during birth, but probably only in certain susceptible individuals [1,2].

Patient

Disease affects a person and induces a process that transforms the healthy subject into a patient – that is, the second component. However, humans are not a homogeneous group – not just with

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2006.02.005

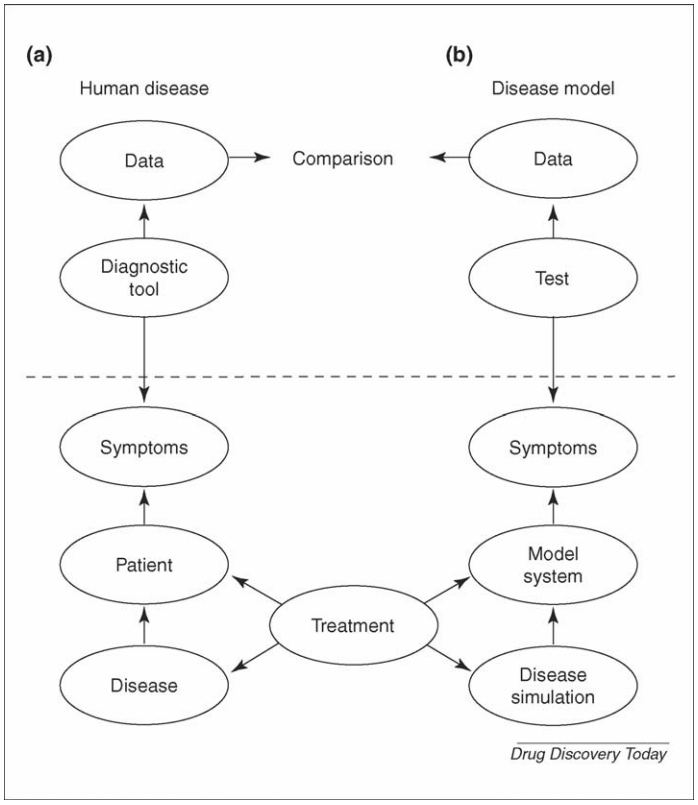

FIGURE 1

Model of the human disease state (a) and the disease model (b).

respect to age and sex, but also in terms of their genetic background. Pharmacogenomic studies have clearly demonstrated that the response to, and metabolism of, drugs show considerable variability between individuals [3–7], and a similar situation is seen for disease processes. For example, in individuals from Guam, it has been observed that the same environmental factor in different individuals can result in the development of AD, Parkinson's disease or amyotrophic lateral sclerosis or a mix thereof, including all of the typical hallmarks of the diseases [8,9]. These observations clearly demonstrate that the genetic background of a person, and possibly their lifetime exposure to environmental factors, can affect the disease pattern and the symptomatology resulting from exposure to a specific disease agent, as well as their response to pharmacological treatment. For the purpose of modelling, it is therefore necessary to consider that a patient and a healthy subject might not be identical in terms of genetic background and physiology.

Symptoms

The disease process in the patient produces a change in the normal physiology of the organism, interfering with the correct functioning of the body and giving rise to a certain symptomatology – that is, clinical features and pathology that we can observe and use to categorize diseases. In recent years, it has become possible to identify the disease-causing agent for some diseases, mostly infectious and parasitic diseases and genetic mutations; based on this information, it has become possible to categorize the disease by cause rather than symptomatology. However, this is rare, and it is, therefore, in most cases necessary to rely on categorization based on symptomatology. Furthermore, we know that different disease-causing agents can give rise to similar symptomatology and that one disease agent can produce different types of symptomatology in different individuals. This means that the disease categories are place holders, which do not necessarily reflect disease mechanisms, and that symptom similarity between two conditions does not prove that the disease-causing mechanisms are identical (Figure 2).

Treatment

Treatment (i.e. the administration of pharmaceutical drugs or manipulations, such as psychotherapy or light therapy) is the only externally controlled component in the system. Treatments can be divided into the categories symptomatic and disease-modifying, depending upon whether they only alleviate specific symptoms or whether they interfere in the disease process itself, hence changing

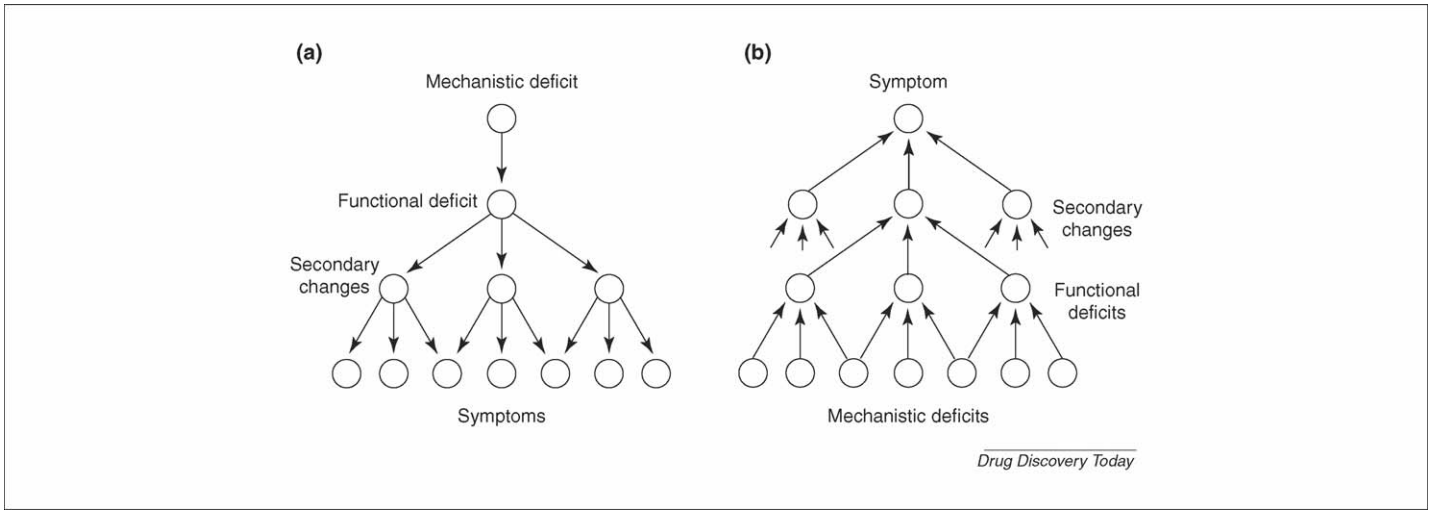

FIGURE 2

Relationship between disease mechanisms and symptoms. (a) The cause of the disease is a perturbation in a certain mechanism, which leads to a set of functional deficits. These give rise to several primary and secondary changes that result in the symptoms that can be observed in the patient. **(b)** A specific symptom can be attributable to several underlying primary and secondary causes, which again might be caused by different functional and mechanistic changes.

the course of the disease. Treatments can also affect non-disease-related processes and can produce treatment-related side effects.

Diagnostic tool

Diagnostic tools are used to evaluate and quantify symptoms, anatomical abnormalities or physiological changes in the patient for the purpose of reaching a diagnosis. They can measure specific symptoms, general physiological parameters, infectious conditions or genetic mutations. When such tools are used as diagnostic criteria for a specific disease, they need to be validated for sensitivity (proportion of people with the target disorder, who have a positive test result) and specificity (proportion of people without the target disorder, who have a negative test result) for that disease [10]. The output of a diagnostic test is a set of data that makes it possible to determine whether the patient is within the normal range, whether a disease state could be present and, in a patient diagnosed with the disease, whether the patient is responding to treatment. Each diagnostic tool normally measures a single parameter, and for the diagnosis of a disease it is frequently necessary to apply several diagnostic tools to differentiate it from other diseases with overlapping characteristics.

The disease model

The disease model (Figure 1b) should, as accurately as possible, reproduce the clinical disease condition, and it can therefore be best described by a model that contains similar components to the clinical model: that is, the model system corresponds to 'patient'; disease simulation to 'disease'; symptoms to 'clinical symptoms'; treatment to 'clinical treatment'; and test to 'diagnostic tool'.

Model system (organism and complexity)

The model system or the preparation is described by two parameters: origin and complexity. The origin of the preparation reflects the organism or species in which the disease is modelled and it can range from bacteria used in an Ames test to the patient, including all the intermediate steps – for example, from *Caenorhabditis elegans* to Zebra fish to rodents to primates to healthy volunteers. The complexity of the preparation reflects the degrees of interactions between the components that can be studied (e.g. increasing the complexity from one cell to two cells will enable the study of cell–cell interactions). It can range from a component in a cell (e.g. a receptor) to a cellular organelle (e.g. the mitochondria) to a cell, to two cells, to a cell cluster, to a tissue, to an entire organ and, finally, to an intact organism.

Disease simulation

In many disease models, it will be necessary to reproduce or mimic the clinical disease state. This method has been described using many different names (e.g. disease-inducing agent and model compound) but here I have used the term 'disease simulation'. Simulations are defined as: 'attempts to represent certain features of the behaviour of a physical or abstract system by the behaviour of another system' (http://en.wikipedia.org/wiki/Simulation). Table 1 lists examples of disease simulators.

Symptoms

The disease simulation will, in the model system, induce a certain symptomatology, which can be compared with the symptomatology observed in the human patient. In some cases, the resemblance can be very close; for example, 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP), a neurotoxin that kills dopaminergic neurones, can induce symptoms in humans and monkeys that

TABLE 1

Categorization of disease simulators

Category	Description
Genetic	Transfer of disease-associated genes into a model system
Mutations	Genetic mutation (e.g. familial forms of Alzheimer's disease)
Risk factors	Genetic disease risk factors, which under certain environmental conditions increase the risks of developing a certain disease (e.g. depression or type II diabetes)
Simulated	External factors which the experimenter applies to reproduce or mimic the disease state in the model system
Physical	These can include procedures to induce a stroke episode, partial nerve ligation to mimic neuropathic pain or manipulating the environment to induce a certain behavioural state (e.g. stress exposure to mimic depression)
Toxins	For example, MPTP, which destroys dopaminergic neurones as a method of mimicking the cell death of these neurones in Parkinson's disease patients
Drugs	Several drugs can produce physiological and behavioural effects that resemble certain diseases (e.g. d-amphetamine and phencyclidine can induce a condition resembling schizophrenia)
Genetic	In some conditions, gene expression levels are changed and these can be mimicked in Tg mice to simulate specific components of the disease (e.g. large T antigen for studying tumour diseases)
Analogous	Several animal species naturally express conditions resembling human disease states (e.g. epilepsy and diabetes)
Environmental	Infectious agents (e.g. viral infections, inducing a concussion)
Intrinsic	Cell, tissue, organ and plasma samples can be obtained from patients, and this material contains the true disease state and the correct genetic background for expressing the disease
Side effects	Some drug-related side effects are primarily seen in patients because the side effects are related to the disease process; however, in many cases, similar side effects would also be observed in healthy volunteers, indicating that the side effects are only drug- or mechanism related. For these latter cases, it is relevant to use healthy animals or material from healthy volunteers

closely resemble the symptoms seen in Parkinson's disease patients, whereas the resemblance in a species such as *C. elegans* is much less obvious, even if MPTP has the same effects [11–13]. To the extent that clinical information is available, symptom comparison does not have to be limited to behavioural or physiological changes but can also include pathway analysis at the level of cells, molecular characteristics and so forth.

Treatment

The evaluation of treatments in the disease model makes it possible to determine the degree to which a disease model reproduces the clinical observations; the better the correlation, the higher the probability that the disease model accurately reflects certain aspects of the clinical condition. However, to perform such a comparison, it is necessary that the treatment regimen used in the disease model mimics the clinical situation as closely as possible because in the case of a poor correlation, it cannot be determined whether the effect seen is due to the disease model or the treatment regimen. The types of treatments that should be tested to evaluate a disease model are drugs known to be clinically effective in the disease, drugs known not to be effective and manipulations that have been shown clinically to affect patients with the disease. Treatment-induced side effects should, for completeness, also be evaluated in disease models; however, in practice, side effects are often evaluated in normal animals because disease models are usually not sufficiently accurate to reflect increased sensitivity to certain side effects in a specific patient population. Because side effects are usually similar in patients and healthy subjects, in this case a normal animal will constitute a suitable model system.

Test

Most of the emphasis in the scientific literature is on the identification of disease-causing factors, whereas relatively little attention is paid to the tests used to evaluate the disease models. However, it is the output of the test that enables comparison with the clinical condition and, for this reason, test selection, data interpretation and influence of confounding factors require as much consideration as the selection of the other components in the disease model.

An example might illustrate this better. The Morris water maze is an excellent test for evaluating spatial orientation and learning. In this test, a rat or a mouse has to learn the location of a hidden platform, which is submerged under water in a water tank measuring 1 or 2 m in diameter. The animal is usually tested over a period of five days, with 3–5 trials per day. Over the course of the training period, the animal learns the external cues around the water tank and, based on these, the location of the platform. On the last day of testing, the animal is given a probe-trial, whereby the platform is removed and the time spent in the area where the platform should have been is used as an index of how well the animal has learned the location of the platform [14,15].

The test is used extensively in disease models of schizophrenia and AD; however, unlike AD patients, schizophrenic patients usually do not suffer from impairments in spatial memory and learning but rather have more-subtle executive function deficits (e.g. difficulty in switching strategy during a cognitive test) [16–21]. The relevance of using the water maze to evaluate cognitive impairments in disease models of schizophrenia is, therefore, questionable because patients typically do not show impairments in spatial navigation.

It is well known that performance in a test situation can be strongly affected by confounding variables such as anxiety, stress, panic responses, motor impairments and sensory impairments; for example, many transgenic mice used for AD research have reduced vision [22]. A parallel example in a human study might involve two young, healthy volunteers being asked to learn the location of a platform in a swimming pool, with one of these volunteers not being able to swim and the other being a competition swimmer. The data would probably show that the competition swimmer had a shorter latency to reach the platform, whereas the other subject spent most of the time along the sides of the pool. For human test subjects, these differences would not be attributed to variations in cognitive ability because it is obvious to us that the nonswimmer has a clear disadvantage. The Morris water maze is used for rats and mice; however, mice are not generally inclined to swim, even though they can, and this means that they are more likely to panic or have an elevated anxiety level, which will have strong effects on the learning curve. Few studies have examined the significance of these confounding factors.

The interpretation of data from a test must also be performed with caution. For example, many studies in the AD field use the probe-trial to measure cognitive performance [23]. However, when studying the swimming pattern of individual animals in detail, it can be seen that some animals continue to swim in the area where the platform used to be, whereas others briefly circle around in this area and then search the rest of the tank looking for the platform. The animals are therefore displaying different search strategies, and both strategies have their advantages and disadvantages. Consequently, it is not possible to label one strategy as being indicative of a cognitive deficit. In the context of AD, animals remaining in the quadrant where the platform used to be are evaluated as having good cognitive performance, whereas in schizophrenia it can be argued that these animals suffer from executive function deficits because they apparently are incapable of switching strategy after realizing that the platform is not where it used to be. This is, of course, absurd but clearly shows that caution must be taken when judging an animal's behaviour.

Several factors must be considered when selecting tests for a disease model: (i) The test must be suitable and biologically relevant for the species being used. (ii) The test must be validated for its specificity and selectivity to measure one or more clinically relevant symptoms or physiological parameters, just as for a clinical diagnostic tool. (iii) Patients can report side effects but animals cannot, and it is therefore necessary to evaluate the possible influence of confounding variables on a result. The doses required to demonstrate behavioural side effects in animals will typically be higher compared with humans because it is necessary to induce side effect levels that interfere with the animal's ability to perform, whereas humans will report a side effect as soon as he or she feels uncomfortable, even though they could still be able to perform normally in a test situation by consciously suppressing the side effect. (iv) Alternative interpretations of test results should be considered which take the natural behaviour of the animal into account, to ensure the objectiveness of any conclusions.

Evaluation and validation of disease models

Disease models show varying degrees of similarity to the human disease state and they might or might not be able to reproduce the clinical effects of drugs to be reproduced. It is, therefore, necessary to develop a terminology and a reference system that enables us to express the level of validity of a disease model and to compare disease models to determine which ones might be most suitable for screening or for the evaluation of a drug.

Face, predictive and construct validity

The system that presently is mostly used was proposed by McKinney and Bunny [24] and later modified by Willner *et al.* [25] and Sams-Dodd [26]. It is primarily applicable to *in vivo* disease models and is routinely used in behavioural pharmacology. It evaluates the validity of a disease model based on three sets of criteria: face validity indicates that the behaviour 'looks like' the clinical condition (i.e. similarity of symptoms); predictive validity indicates that the response to drug treatment and manipulations is similar to that observed clinically; and construct validity was defined by Willner *et al.* [25] to indicate that a model is in accordance with current theories on the disease mechanism, and by Sams-Dodd [26] that the underlying disease mechanism in the model is the same as in the human disease. However, this system has limitations because each component is not evaluated separately, making it difficult to compare very different systems (e.g. *in vitro* versus *in vivo* models), and it does not grade the level of validity, making it difficult to compare disease models.

Disease model validity system

In general, disease models are developed by first selecting the model system and the disease simulation method; second, selecting the test for the clinical parameter to be studied; third, combining disease model and test into a screening system, where the effects of treatments are determined and correlated to clinical findings; and fourth, evaluating the disease model in further tests to determine in how many aspects it resembles the clinical disease state. Combining this with the model in Figure 1, the following relationships can be defined:

Disease model = model system + disease simulation

Where: model system = organism + complexity

Screening system = disease model + test

Model system and disease simulator

The 'model system' (i.e. organism and complexity) and the 'disease simulation' are selected by the experimenter before experimentation and it is therefore possible to evaluate their theoretical level of validity (Table 2). For 'organism', the validity is determined based on phylogenetic proximity to man, and for 'complexity', the degree to which the preparation reflects the full complexity of an intact organism. Disease simulation corresponds to 'disease', and the 'construct validity' reflects the validity of the method for simulating the disease in the model system.

Test

The parameter or symptom that is measured in the disease model is determined by the 'test', and the 'test validity' reflects the degree of correlation between the test and the 'diagnostic tool' used for the diagnosis of this specific symptom or parameter in patients. This is determined by evaluating different disease models in the test (Table 3a). For example, if a test identifies cognitive impairments in disease models of AD, but not schizophrenia, it might indicate that the test measures deficits in spatial learning or working memory but probably not in executive function. This specificity determines the validity of the test for a specific clinical parameter and determines the relevance of using the test in combination with a particular disease model.

TABLE 2

Disease model component validity

	Model system validity		Construct validity
Level	**Organism**	**Complexity**	**Disease simulation**
Not valid			No inducer
Low	Unicellular Invertebrate Vertebrate	Mechanism Cellular organelle Cell	**Simulated**
Medium	**Mammal** Primate Human	Cell cluster Tissue Organ	Analogous Genetic
High	Patient	**Intact organism**	Environmental and/or intrinsic

Before the experimental evaluations, it is possible at a theoretical level to evaluate the clinical validity of the disease model components. The model system corresponds to the patient, and model system validity represents the proximity of the system to the patient (i.e. for organism, phylogenetic proximity to the patient; and for complexity, the degree to which it reflects an intact organism). Disease simulation represents the method of mimicking the disease, and construct validity represents the clinical relevance of the method. To exemplify the use of Tables 2–4, phencyclidine-induced social isolation in rats as a model of the negative symptoms of schizophrenia has been used. Phencyclidine is a drug of abuse, clinically known to mimic the negative symptoms (e.g. social withdrawal and isolation) of schizophrenia in healthy volunteers. The social interaction test measures the level of social behaviour between pairs of rats; in this test, phencyclidine disrupts the social behaviour, and antipsychotic drug treatment can reverse the phencyclidine effects in a manner that correlate to clinical findings. In contrast to phencyclidine, d-amphetamine does not mimic the negative symptoms of schizophrenia in healthy subjects and does not cause disruption of social behaviour in the social interaction test [26]. For Table 2, the validity of the phencyclidine model has been indicated in bold.

Screening system

The 'screening system' is the combination of a disease model with a specific test, and the 'screening system validity' (which corresponds to 'predictive validity') is determined by correlating treatment effects (e.g. drugs) in the screening system to observations in patients (Table 4).

Disease model

'Disease model validity' is determined by evaluating the disease model in several tests, where each test measures a parameter representing a clinical symptom. By correlating the effects of treatments in each of these screening systems to clinical observations (Table 3b), it is possible to determine whether the disease model fully reproduces the clinical syndrome, or can only be used in combination with a specific test to evaluate treatment effects on a specific symptom.

The disease model validity system involves four steps and results in an estimation of the validity for each of the components in the model (Figure 1; Tables 2–4), and this can be used to evaluate and compare tests, screening systems and disease models. In principle, the level of validity for each component should be an absolute

TABLE 3a

Test and disease model validity

	Test (social isolation)	
Disease models	**Test validity**	**Comments**
Phencyclidine model of schizophrenia	Correct	Phencyclidine induces social isolation
Amphetamine model of schizophrenia	Correct	Amphetamine does not affect social behaviour
Aged animals	Correct	Aged rats show reduced social behaviour

Test validity is determined by applying the test to several different disease models to validate that the test accurately measures a specific symptom or parameter.

TABLE 3b

Test and disease model validity

	Phencyclidine model of schizophrenia	
Tests	**Model validity**	**Comments**
Social isolation	Yes – medium	Induction of social isolation + treatment effects
Prepulse inhibition	Yes – medium	Disruption of prepulse inhibition + treatment effects
Water maze	Yes – high	No effects of phencyclidine

Disease model validity reflects the ability of the disease model comprehensively to reproduce the clinical syndrome and is determined by evaluating the disease model in different tests and correlating these findings to clinical observations. In the evaluation, the validity of the different disease models and tests should be considered to avoid the results from an inferior disease model or test being weighted equally with a highly valid model or test. For the social interaction test, findings show that phencyclidine, but not d-amphetamine, disrupt the social behaviour among pairs of rats, indicating that the test can measure behaviours corresponding to negative symptoms. However, aged animals also show reduced social interaction, indicating that the test is not specific for negative symptoms but for deficits in social interaction. For the phencyclidine model, the disease model table (Table 3a) shows that phencyclidine, as expected based on clinical observations, causes deficits in social interaction and prepulse inhibition (a test for sensory information processing), and that it does not cause spatial navigation deficits (based on experiments in the author's laboratory). Together, these findings suggest that phencyclidine in rats might be a valid symptomatic model of schizophrenia.

level relative to the clinical condition; however, in practice it is more realistic when comparing disease models and tests to determine the level of validity of the components relative to each other because the clinical conditions are often not well defined, and the diagnostic tools could be based on parameters that cannot be evaluated in animals (e.g. interview-based rating scales). Finally, in case the disease model does not exhibit a high degree of validity, it is likely that either the model system or the disease simulation is not valid, and it will therefore be necessary to modify these components. This is a time-consuming process and for this reason the highest theoretical level of validity to the clinical situation should be sought initially.

TABLE 4

Screening system validity

	Screening system validity
Level	**Clinical correlation**
Not valid	No correlation
Low	**Drug class selectivity** **Onset of action**
Medium	Relative efficacy Nonresponders
High	Identical

The screening system is the combination of the disease model with a specific test. The validity of the system is determined experimentally by the testing of drugs and other manipulations and by correlating the findings to clinical observations. Screening system validity (also called predictive validity) therefore represents the degree to which treatment effects in the screening system correlate to clinical findings for one specific clinical parameter. The levels for the phencyclidine-induced social isolation in rats are indicated in bold [26].

There is an important relationship between the validation of a test and a disease model, which should be noted. A disease model is validated by applying different tests to it, whereas a test is validated by evaluating it in different disease models. However, this means that the disease model and test are validated against each other without knowing whether in reality either reflects the clinical condition. It is only possible to avoid these circular arguments in disease models based on environmental or intrinsic disease simulations.

Drug screening systems

In Table 5, model system and disease simulation validity have been evaluated for six commonly used screening systems, in terms of the number of components differing from the clinical state. The receptor-based screening system normally uses a cell-free system or cells from a mammal, it has the simplest level of complexity (i.e. a receptor) and there is not a disease simulator to mimic the disease state. This means that all three columns differ from the clinical situation. At the other end of the spectrum is the transgenic animal model, in which a familial gene mutation has been transferred to

TABLE 5

Drug screening systems

	Organism	Complexity	Disease simulation	Discrepancies to clinic
Receptor-binding assay	Mammal	Mechanism	None	3
Animal tissue	Mammal	Tissue	Artificial	3
Isolated organ	Mammal	Organ	Artificial	3
Animal disease model	Mammal	Intact organism	Artificial	2
Patient tissue	Patient	Tissue	True disease	1
Tg animal	Mammal	Intact organism	True disease (familial form)	1 – familial, 2 – sporadic
Human patient	Patient	Intact organism	True disease	0

The levels of disease model validity for six commonly used drug screening systems are compared, based on the number of components differing from the clinical state (discrepancies). The human patient has been included for comparison.

the animal. The system is based on a mammal, the system is intact and the disease simulator corresponds to the true disease state because it is based on a familial mutation (i.e. the number of columns differing from the clinical state is one). However, the number is two if the disease model is used to evaluate treatments for sporadic forms of the disease because the mechanism in sporadic forms might be different (a good example is leptin as a treatment for obesity [27,28]). The most surprising conclusion from Table 5 is the relatively high level of validity for the use of patient material. For this disease model, the 'organism' and the 'disease' are the clinical condition, and it is only the complexity of the preparation that differs from the clinical state. Patient material is rarely used in drug discovery, in spite of the fact that the genetic background and the disease state are intrinsic to the model, thus avoiding the difficult task of having to simulate a disease when the cause of the disease is unknown.

An important aspect to consider when selecting a drug screening system is whether the goal is to develop a symptomatic versus a disease-modifying treatment. For infectious diseases, the disease cause is usually known and this has enabled the development of 'cures' – that is, treatments that target the cause of the disease. However, for most noninfectious diseases, the underlying disease mechanism is unknown, with the result that disease models can only be developed and validated based on symptom-similarity to the clinical condition, and this might explain why most treatments for noninfectious diseases only are symptom reducing and not disease modifying [28,29]. This relationship is shown schematically in Figure 2. In Figure 2a, it is assumed that a disease causes a deficit or a change in a mechanism, and this results in an abnormality in the normal functioning of the cell (e.g. axonal transport, mitochondrial function or calcium storage). This functional abnormality affects cell function, and this will affect the normal functioning of the organism and give rise to the wide range of symptoms that can be observed in a patient. This is the normal pattern for a disease; however, in drug discovery it is often necessary to develop a disease model based on symptom similarity, without understanding the true disease mechanism, and, as shown in Figure 2b, these models will rarely be valid for identifying disease-modifying treatments. The reason for this is that any specific symptom can be the result of several different underlying functional changes, and each of these might again be attributed to an even wider array of deficits at the mechanistic level. The probability of selecting the right disease mechanism when developing a disease model for a symptom is therefore very low, with the consequence that treatments developed based on such a disease model most probably will only alleviate symptoms and will not be disease modifying. In this context, an interesting observation is that studies at the level of functional changes in cells and tissue in AD have shown that identical changes can be seen at this level without being reflected at the higher symptomatic level; for example, neurones from AD patients and Tg mice carrying a human mutation causing AD show the same deficits in axonal transport mechanisms, yet the mice do not show prominent cognitive deficits [30,31]. These observations suggest that a human mutation in a mouse can cause the same mechanistic and functional changes as in the patient at the cell and tissue level but, because of species differences, the overall symptomatic consequences are different [32]. This could suggest that a more successful approach for the identification of disease-modifying treatments might be to study functional changes at the level of cells and tissue from patients with the disease or from Tg mice carrying a human mutation.

The proof-of-principle study

The goal of the proof-of-principle study is to demonstrate in a well-validated disease model that a compound has the expected *in vivo* efficacy before proceeding into clinical testing [33]. For this purpose, the study must mimic a clinical drug trial as closely as possible, to achieve the highest possible level of predictive validity. For a clinical trial, the first step is to define inclusion and exclusion criteria for patients – that is, which diagnostic subtypes of patients can be included and which types of co-morbidity represent exclusion criteria (e.g. epilepsy, a heart condition, AD, etc.). The second step is to define the clinical scales that will be used to evaluate the patients and the types of side effects that must be recorded, either by physiological readouts (e.g. electrocardiogram) or interviews (e.g. nausea or dizziness). The third step is to initiate the study using a drug administration regimen that is adapted to the pharmacokinetic properties of the compound (e.g. once- or twice-daily administration), and a dose regimen that stays below the maximum tolerated dose determined in the Phase I study. The duration of the study depends upon the indication because some drugs might have a delayed onset of action, and it can take 1–2 years to see the treatment effects (e.g. disease modification in AD). The fourth step is to evaluate whether the compound has superior therapeutic efficacy relative to marketed drugs and to ensure that the study was conducted appropriately, so a comparator compound is included.

The goal is to transfer the clinical trial situation as accurately as possible to the proof-of-principle study in an *in vivo* disease model. It is assumed that the chosen disease model has previously been validated and that tests have been identified that can measure the relevant parameters in this model. The first step is to consider whether the included animals should be preselected to reduce variability in the study. This will depend on the specific model being used. The second step is to decide which tests to include. These tests should reflect the parameters being measured in the clinic as closely as possible, to maximize the predictive validity. The evaluation of side effects is usually performed in separate treatment groups but they should include parameters that might interfere with the primary effect parameters – that is, confounding variables, as well as side effects, that might be important to enable a decision to be made as to whether one should proceed with a compound (e.g. sedative properties and interference with cognitive functions). The third step is to decide on the drug administration regimen, which should mimic the clinic as closely as possible. Suitable biomarkers should, if possible, be incorporated into the study design to monitor the biological effects of the drug. With respect to dose selection, this can often be established based upon results from previous studies, from which the active dose range typically will be known, but in principle the study design should include the construction of a full dose–response curve. This information is necessary to determine the therapeutic index relative to serious side effects and to determine the shape of the dose–response curve. For example, a narrow U-shaped dose–response curve might represent a huge problem for a doctor treating a

patient because, in the absence of an effect or a suitable biomarker, how could one determine whether the dose should be increased or decreased? The administration regimen should include at least one chronic dosing study (e.g. 4–6 weeks) to ensure that tolerance or sensitization effects do not develop to the compound. For the fourth step, a comparator compound should be included to demonstrate that the experiment has functioned as expected and to position the new compound with respect to existing therapy – that is, does it represent a therapeutic advantage over existing compounds? Finally, at the end of the experiment, plasma and target tissue levels of the compound should be determined to ensure that drug exposure had been sufficient.

It might be necessary to use a study design that is more rapid than the optimal solution (e.g. during a lead optimization phase, where compounds have to be tested on a regular basis) but the faster design should be validated against the full design to avoid assumption errors. For example, in stroke research, compounds were selected preclinically based on their ability to reduce neuronal death using compound administration at the time of infarct, and the neuronal survival was measured 24 h after the infarct. However, in the clinic, stroke compounds can rarely be administered earlier than 4–5 h after the infarct, and the outcome parameter is usually functional recovery 2–3 months after the infarct [34–36]. Clinical trials found the drugs to be ineffective, and subsequent experiments have shown that the drugs had similar effects in animals when tested under conditions closely mimicking those seen in the clinic.

The proof-of-principle studies can also provide important information on the mode of action of the compound. From the experiments, a biologically active dose can be determined, as well as the corresponding drug concentration in the target tissue. Using this drug concentration, together with a broad receptor, ion channel or enzyme screen, it is possible to identify, based on the IC_{50}s of the drug at different receptors, which receptors could have contributed to the biological effect at the active concentration. Although most drug programmes are aimed at developing highly selective compounds, when the drugs are tested *in vivo*, the effective concentration often does not fit with the nanomolar selectivity that can be obtained for the compound in, for example, a receptor-binding assay. The mapping of the activated mechanisms at the biologically-active concentration in the target tissue can, therefore, be helpful to identify the true mechanism(s) of action.

Conclusion

The purpose of this review has been to provide a theoretical discussion of the concepts underpinning disease models and to develop a terminology and a framework to analyze and express the clinical validity of disease models. The advantages of the proposed framework, which builds on the concepts of McKinney and Bunney [24], is that it can be used for *in vitro* and *in vivo* systems and that it enables the comparison of disease models to determine their relative level of validity.

The proposed framework compares and evaluates disease models based on their validity relative to the clinical situation, and the key conclusion of the paper is the necessity of integrating clinical information fully into the process of developing disease models, into the selection of tests to be used with a disease model and into the process of evaluating and comparing the validity of different disease models. We have limited understanding of the biology of the organism and of diseases, and this means that our only real yardstick is clinical information. This must consequently be used to the fullest extent to identify the assumptions that are built into disease models because only by identifying these assumptions will we become able to examine those that are testable, to increase the validity of our model. There will always be assumptions that cannot be tested and which we will be forced to accept until we have a better understanding of the biology of diseases; however, a systematic approach to this process will improve our knowledge of the strengths and limitations of a disease model and this will help us to avoid making similar mistakes to those made in the past (e.g. the stroke programmes). This will mean making concrete financial savings in terms of avoiding unnecessary clinical trials and faster termination of drug programmes that, in any case, are destined for failure; more importantly, an improvement of our disease models might also lead to the discovery of new targets and approaches to disease treatment and could ensure their evaluation in clinical trials.

Acknowledgements

I thank Jeanette Sams-Dodd for commenting on the manuscript.

References

1 Hirsh, S.R. and Weinberger, D.R., eds) (1995) *Schizophrenia*, Blackwell Science Ltd.
2 Boog, G. (2004) Obstetrical complications and subsequent schizophrenia in adolescent and young adult offsprings: is there a relationship? *Eur. J. Obstet. Gynecol. Reprod. Biol.* 114, 130–136
3 Kassam, S. *et al.* (2005) Single nucleotide polymorphisms (SNPs): history, biotechnological outlook and practical applications. *Curr. Pharmacogenom.* 3, 237–245
4 Wechsler, M.E. and Israel, E. (2005) How pharmacogenomics will play a role in the management of asthma. *Am. J. Respir. Crit. Care Med.* 172, 12–18
5 Penny, M.A. and McHale, D. (2005) Pharmacogenomics and the drug discovery pipeline: when should it be implemented? *Am. J. Pharmacogenom.* 5, 53–62
6 Phillips, K.A. and Van Bebber, S.L. (2005) Measuring the value of pharmacogenomics. *Nat. Rev. Drug Discov.* 4, 500–509
7 Lewin, D.A. and Weiner, M.P. (2004) Molecular biomarkers in drug development. *Drug Discov. Today* 9, 976–983
8 Morris, H.R. *et al.* (2004) Genome-wide analysis of the parkinsonism-dementia complex of Guam. *Arch. Neurol.* 61, 1889–1897
9 Murch, S.J. *et al.* (2004) A mechanism for slow release of biomagnified cyanobacterial neurotoxins and neurodegenerative disease in Guam. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12228–12231
10 Whiting, P. *et al.* (2004) Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technology Assessment* 8 no. 25
11 Smeyne, R.J. and Jackson-Lewis, V. (2005) The MPTP model of Parkinson's disease. *Brain Res. Mol. Brain Res.* 134, 57–66
12 Jenner, P. (2003) The MPTP-treated primate as a model of motor complications in PD: primate model of motor complications.. *Neurology* 61 (6, Suppl. 3), S4–S11
13 Braungart, E. *et al.* (2004) *Caenorhabditis elegans* MPP+ model of Parkinson's disease for high-throughput drug screenings. *Neurodegenerative Dis.* 1, 175–183
14 Morris, R.G.M. *et al.* (1982) Place navigation impaired in rats with hippocampal lesions. *Nature* 297, 681–683
15 De Hoz, L. *et al.* (2005) Spatial learning with unilateral and bilateral hippocampal networks. *Eur. J. Neurosci.* 22, 745–754
16 Behl, P. *et al.* (2005) Progress in clinical neurosciences: cognitive markers of progression in Alzheimer's disease. *Can. J. Neurol. Sci.* 32, 140–151

17 Merchant, C. and Hope, K.W. (2004) The quality of life in Alzheimer's disease scale: direct assessment of people with cognitive impairment. *J. Clin. Nurs.* 13, 105–110
18 Schmitt, F.A. *et al.* (2002) Measuring cognition in advanced Alzheimer's disease for clinical trials. *J. Neural Transm. Suppl.* 62, 135–148
19 Bowie, C.R. and Harvey, P.D. (2005) Cognition in schizophrenia: impairments, determinants, and functional importance. *Psychiatr. Clin. North Am.* 28, 613–633
20 Braff, D.L. and Light, G.A. (2004) Preattentional and attentional cognitive deficits as targets for treating schizophrenia. *Psychopharmacology (Berl.)* 174, 75–85
21 Goldberg, E. and Bougakov, D. (2005) Neuropsychologic assessment of frontal lobe dysfunction. *Psychiatr. Clin. North Am.* 28, 567–580
22 Li, G. *et al.* (2003) β-amyloid secretases and β-amyloid degrading enzyme expression in lens. *Mol. Vis.* 9, 179–183
23 Janas, A.M. *et al.* (2005) The cholinesterase inhibitor, phenserine, improves Morris water maze performance of scopolamine-treated rats. *Life Sci.* 76, 1073–1081
24 McKinney, W.T., Jr and Bunney, W.E., Jr (1969) Animal model of depression. I. Review of evidence: implications for research. *Arch. Gen. Psychiatry* 21, 240–248
25 Willner, P. *et al.* (1992) Chronic mild stress-induced anhedonia: a realistic animal model of depression. *Neurosci. Biobehav. Rev.* 16, 525–534
26 Sams-Dodd, F. (1999) Phencyclidine in the social interaction test: an animal model of schizophrenia with face and predictive validity. *Rev. Neurosci.* 10, 59–90
27 Yanovski, S.Z. and Yanovski, J.A. (2002) Obesity. *N. Engl. J. Med.* 346, 591–602
28 Veniant, M.M. and LeBel, C.P. (2003) Leptin: from animals to humans. *Curr. Pharm. Des.* 9, 811–818
29 Sams-Dodd, F. (1998) A test of the predictive validity of animal models of schizophrenia based on phencyclidine and d-amphetamine. *Neuropsychopharmacology* 18, 293–304
30 Stokin, G.B. *et al.* (2005) Axonopathy and transport deficits early in the pathogenesis of Alzheimer's disease. *Science* 307, 1282–1288
31 Billings, L.M. *et al.* (2005) Intraneuronal Aβ causes the onset of early Alzheimer's disease-related cognitive deficits in transgenic mice. *Neuron* 45, 675–688
32 Colton, C.A. *et al.* (2005) NO2 regulation of tau hyperphosphorylation in a mouse model of AD. In *7th International Conference AD/PD 2005*:108.
33 Sams-Dodd, F. (2005) Target-based drug discovery: Is something wrong? *Drug Discov. Today* 10, 139–147
34 Legos, J.J. *et al.* (2002) Pharmacological interventions for stroke: failures and futures. *Expert Opin. Investig. Drugs* 11, 603–614
35 Stroke Therapy Academic Industry Roundtable (STAIR) (1999) Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke* 30, 2752–2758
36 Calkins, K. and Wess, L. (2003) Stroke: never say die. *BioCentury* 11, A1–A7