ELSEVIER

Improving the predictive value of interventional animal models data

Caroline J. Zeiss

Section of Comparative Medicine, Yale University School of Medicine, 375 Congress Ave, New Haven, CT 06520, USA

For many chronic diseases, translational success using the animal model paradigm has reached an impasse. Using Alzheimer's disease as an example, this review employs a networks-based method to assess repeatability of outcomes across species, by intervention and mechanism. Over 75% of animal studies reported an improved outcome. Strain background was a significant potential confounder. Five percent of interventions had been tested across animals and humans, or examined across three or more animal models. Positive outcomes across species emerged for donepezil, memantine and exercise. Repeatable positive outcomes in animals were identified for the amyloid hypothesis and three additional mechanisms. This approach supports *in silico* reduction of positive outcomes bias in animal studies.

Introduction

The translation of basic biomedical knowledge into effective treatment for human disease has relied heavily on the use of animals as models. However, for many complex disorders, therapeutic success in animals has not been accompanied by similar success in humans [1–3]. One increasingly cited reason lies in flawed animal study design [4–7] and insufficient reporting of methods [4]. These methodologic issues probably contribute to failure of repeatability and efficacy within and across the animal–human boundary [8–10].

The reasons for poor translatability of animal data are multifactorial [1]. Apart from demonstrating efficacy in a model system, pharmacokinetic and pharmacodynamic (PK/PD) variables that define the target exposure–response relationship for a given intervention across species must be established [11]. Recent methodologic developments in PK/PD modeling [11,12], pathway-based toxicology [13,14] and emerging multiorgan and 3D tissue culture technologies [15] have the dual benefit of improving predictability of these processes and reducing animal use [16]. Regardless of improvements in these areas, efficacy must still be established – this is likely to continue to drive animal model use for the foreseeable future, despite suggestions that the animal model paradigm is broken [17]. Published literature represents a valuable source of efficacy data, with the caveat that there are some challenges to its interpretation. The current research landscape is one in which the majority of interventional animal studies report improved outcomes [2,18–21]. This trend derives from the combination of reluctance to report negative data [19], and methodologic flaws that promote false-positive outcomes [20,21]. One way to assess the therapeutic potential of an intervention is to examine the methodologic details of related studies across species and settings [22]. To conduct such a meta-analysis, related studies assessing the same intervention must first be aggregated. For complex diseases driven by multiple intersecting mechanisms [23], it is a challenge to aggregate the evidence for or against the therapeutic potential of a given mechanism or intervention within the full spectrum of current work on the topic. This is further complicated by the increasing number of studies published each year [24], with the attendant risk that many useful animal studies are not identified. This review approaches this challenge by assessing repeatability of interventional outcomes across mechanisms and species, using a networks-based systems approach. For this proof-of-concept review, Alzheimer's disease (AD) is used as an illustrative example of a complex condition.

Methods

448 interventions across 752 animal and human studies and human clinical trials were examined. Briefly, source data were aggregated from PubMed, Mouse Genome Informatics (MGI) and

E-mail address: caroline.zeiss@yale.edu.

1359-6446/06/$ - see front matter http://dx.doi.org/10.1016/j.drudis.2014.10.015

GLOSSARY

Intervention Introduction of variable (pharmaceutical, phytochemical, physical, genetic, behavioral or environmental) into a model system with the intent to assess its effect upon outcome.
Mechanism Means by which an intervention influences a biological system such as a mechanism of action of a drug, effect on a cellular pathway or a physiologic function.
Model An animal system harboring specific AD-related alleles (e.g. APP E618Q), or an animal in which some mechanistic aspect of AD is modeled (e.g. intracerebroventricular streptozotocin injection in wild-type animals).
Outcome Effect of an intervention on overall disease severity in a patient population or animal model system, as defined by the authors in the abstract.
Patient population These constituted study populations with diagnosis of late-onset Alzheimer's disease, familial Azheimer's disease or normal aging, as defined in publications and in ClinicalTrials.gov.
Strain background The rodent strain upon which the mutant allele resides (e.g. C57BL/6J).

ClinicalTrials.gov using search terms and time limits listed (see supplementary material Table S1 online). Only intervention studies were included, reducing the number of studies to a total of 353 and 148 interventional animal and human studies respectively, and 251 human clinical trials (see supplementary material Table S1 online). Intervention studies were defined as those in which the effect of a broadly defined intervention (pharmaceutical, phytochemical, physical, genetic, behavioral or environmental) on the AD phenotype was examined (see Glossary). Next, the following information was collected for each clinical trial or study: patient population (human studies), model (for animal studies), strain background (for rodent studies), intervention, outcome and mechanism. Outcomes described the effect of an intervention on overall disease severity in a patient population or animal model system (see Glossary). In the majority of human studies, outcomes were based on clinical measures, whereas in animal studies a range of biomarkers, pathologic scores and/or functional measures were used. Each study was also assigned a mechanism based on the mechanism of action of its intervention. To establish a controlled vocabulary across species, mechanisms were defined by GO (Gene Ontology; http://geneontology.org/) terms. Related interventions were aggregated within their cognate mechanisms, thus providing a central framework around which to organize interventional studies. Individual data segments (patient population, model, intervention, outcome and mechanism) were arranged in binary form, to denote whether two entities (nodes) had a relationship (edge) or not, without imposing a preferential value on any relationship. To visualize the network, a freely available network program, Biolayout Express 3D 3.0 (http://www.biolayout.org/) [25] was used. Detailed methods used to aggregate the dataset for the review (Table S1), lists of search terms (Table S2), results of searches (Tables S3–S6; Figures S1 and S2) are presented in the supplementary material.

Results

Utility of the network

The simplest use of the network is to rapidly compile animal and human outcomes data around a given intervention (supplementary material Figure S1), as well as to identify studies across related interventions that map to a common mechanism (Figure S2; Table S6a,b). Additionally, for the purposes of this review, the network was interrogated to explore patterns regarding translatability of animal model data in AD. Four questions were posed: (i) Can patterns of animal model use be identified? (ii) How do outcomes segregate across species? (iii) What mechanisms have been studied across species? (iv) Can interventions and mechanisms with translational potential be identified? For the purposes of this review, interventions with translational potential were defined as those in which evidence was provided for similar outcomes across multiple species and settings.

Can patterns of animal model use be identified?

Several findings emerged that are likely to contribute to failure of promising preclinical candidates to translate in human AD trials. Of the 139 models across 11 species included in the network only 20 were used in four or more studies, and one model [Tg(APPswe,PSEN1dE9)85Dbo] was used in 24% of studies (Table 1). This is likely to represent a form of publication bias, in which well-accepted models engender continued use. The most heavily used strains, with the human AD alleles they model, and their strain background (as designated by MGI, see Methods online) are listed in Table 1.

Precise definition of rodent models was rarely provided. Mouse models were often identified using synonyms (e.g. the synonym APP/PS1 denotes four different models; Table 1). Model nomenclature defining the mutant allele was usually provided in the methods or references; however background strain was rarely described. Because precise strain nomenclature was incorporated in the dataset, the presence of potential genetic confounders across all genetically altered mouse studies could be identified. Of the 16 strains in Table 1 (comprising approximately 80% of animal studies in the network), eight contain contributions from strains carrying the *rd-1* allele. This autosomal recessive allele results in photoreceptor degeneration and renders animals blind by 6–8 weeks of age [26]. Because its presence cannot be identified clinically, the *rd-1* allele is propagated in models on mixed backgrounds, thus rendering an unknown proportion of mutant and control populations blind. Inconsistent results and high within-group variability of mutant and wild-type groups in spatial tests used to detect memory impairment have been noted [27–29]. One commonly used test, the Morris water maze, has been shown to be strongly impacted by presence of the *rd-1* mutation associated with the SJL background [30] in a commonly used model [Table 1; APP695swe: Tg(APPSWE)2576Kha]. This model is also available on the 129S6 background strain (http://www.taconic.com/2789), which carries the *Disc1* mutation known to affect working memory in mice [31].

A further three models in Table 1 contain contributions from the DBA/2 strain. DBA/2J mice harbor two mutations that result in progressive eye abnormalities that closely mimic human hereditary glaucoma [32]. In total, 55% of the interventional studies using AD models in 2013 were done in populations potentially carrying mutations (at unknown frequencies) that impair vision. The potential impact of confounding alleles is compounded by the tendency for animal studies to use small sample sizes [10,33], resulting in overestimation of effect size, high false-positive rates