Published in final edited form as:

Drug Discov Today. 2015 April ; 20(4): 475–482. doi:10.1016/j.drudis.2014.10.015.

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

Abstract

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.

Keywords

Translational; animal; network; intervention; preclinical; trials

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 methodology issues probably contribute to failure of repeatability and efficacy within and across the animal–human boundary [8–10].

© 2014 Elsevier Ltd. All rights reserved.

Corresponding author: Zeiss, C.J. (caroline.zeiss@yale.edu).

Teaser: Using Alzheimer's disease as a case study, 448 interventions across species were interrogated using a networks-based method to assess repeatability of outcomes by intervention and mechanism.

Author contribution

Caroline Zeiss aggregated and curated source data, established network algorithms, performed all subsequent analyses and drafted the manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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 (see Glossary) across species must be established [11]. Recent methodology 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 [2,19–22]. This trend derives from the combination of reluctance to report negative data [20], and methodology flaws that promote false-positive outcomes [21,22]. One way to assess the therapeutic potential of an intervention is to examine the methodology details of related studies across species and settings [23]. To conduct such a meta-analysis, related studies assessing the same intervention must first be aggregated. For complex diseases driven by multiple intersecting mechanisms [24], 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 [25], 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

Across 752 animal and human studies and human clinical trials 448 interventions were examined. Briefly, source data were aggregated from PubMed, Mouse Genome Informatics (MGI) and 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. Next, the following information was collected for each clinical trial or study: patient population (human studies), model (for animal studies), intervention, outcome and mechanism. Outcomes described the effect of an intervention on overall disease severity in a patient population or animal model system. 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 the gene ontology project 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/>) 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,S2) are presented in the supplementary material.

Results

Utility of the network

The simplest use of the network is rapidly compiled 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(APP^{swe}, PSEN1^{dE9})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 supplementary material 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(APP695swe)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 [7] and poor reproducibility. Studies could indicate that the parent strain has been backcrossed to a single strain, usually C57BL/6, and not all studies employ cognition testing that could be impacted by blindness. However, if this is not noted, or the use of F1 hybrids (without subsequent breeding) between *rd-1* and non-*rd-1* carrying strains is not explicitly stated, the distribution of blinding alleles in mutant and control populations cannot be accurately assessed.

Third, variations in species physiology that impact translatability were identified. These included descriptions of strain variation in processing of amyloid-beta [34], altered phenotype expression by strain [35–37], variations in behavioral testing by strain [38,39], species similarities [40] or differences in molecular signatures of the aging synapse [41], or aging brain [42,43], gender differences in amyloid plaque development within model [43,44] and differences in amyloid-beta degradation and toxicity between mice, humans and large animal models [45,46]. This type of normative data profoundly impacts study design and interpretation, and contributes to the species barrier [18]. Unfortunately, it is difficult to find without prior knowledge of what to look for.

How do outcomes segregate across species?—Interventional outcomes across humans and animal studies revealed a translational gap (Figure 1; Table 2). The majority of animal studies in 2013 (75.9%) reported an improved outcome. Just under half of published human studies (46.6%) over a five year period reported improvement. It should be noted that the denominator of the human data included published reports only (Table 2), and thus excluded the majority of human clinical trials. The majority of human clinical trials (89%) have no results available through the ClinicalTrials.gov site. Animal studies with a worsened outcome (13.8%) were predominantly those where the intent was to worsen some aspect of the AD-like phenotype (e.g. cholinergic denervation) [47]. Therefore, these are not directly comparable with the few human studies in which an intervention intended to treat the disease actually worsened outcome [48–51].

What mechanisms have been studied across species?—Animal and human studies mapped to a total of 81 different mechanisms (supplementary material Table S3), implicating a process of cumulative decline across multiple cellular systems in AD. Of these, 42 mechanisms were shared across humans and animals (i.e. they contained at least one intervention tested in humans and animals). The proportion of studies assigned to these

common mechanisms differed across species (Figure 2; supplementary material Table S3). Studies targeting neurotransmitter systems predominated in humans (constituting 38.8% of all interventions in humans compared with 10% in animals). This reflects interventions across multiple neurotransmitter systems for psychological manifestations of AD such as apathy, depression, aggression and agitation [52–54]. Within this category, studies targeting the cholinergic system (GO:0015464 – acetylcholine receptor activity) predominated in humans and animals. Interventions that damage the cholinergic system in mice [47,55] and monkeys [56,57] promote amyloidogenic damage and impair cognition, thus providing evidence for intersecting cholinergic and amyloid hypotheses. Studies focused on amyloid-beta-related interventions were similarly distributed across species (21.8% and 18.5% of interventions in humans and animals, respectively; blue shading in Figure 2). Studies examining mechanisms that can be ascribed to lifestyle changes of diet, exercise or nutraceutical supplementation (GO:0042593 – glucose homeostasis; GO:0014823 – response to exercise; GO:0006766 – vitamin metabolic process; GO:0007584 – response to nutrients) assume almost identical proportions across humans and animals (11.2 and 11.6%, respectively). Sixty-four percent of 25 human studies across these four mechanisms demonstrate improved outcomes. Of the 30 studies examining the outcomes of three approved interventions that map to cholinergic mechanisms (GO:0015464 – acetylcholine receptor activity – donepezil, rivastigmine and galantamine), 20 (66%) had improved outcomes. Interventions rooted in inflammation (red box in Figure 2), oxidative stress (gray box in Figure 2) and glucose homeostasis (purple box in Figure 2) prevailed in animals and reflect the relatively greater ease with which biomarkers for these mechanisms can be quantified in the brains of animals compared with humans.

Can interventions and mechanisms with translational potential be identified?

—The large number of positive outcomes in animal studies presents an interpretive challenge. To address this, a three-step approach was used. First, (i) the network was interrogated to identify characteristics of animal studies for those interventions of known benefit in humans. Next, (ii) the repeatability of outcomes for those interventions that had been examined across multiple animal models or species was assessed. Last, (iii) the repeatability of outcomes across multiple animal models or species for those mechanisms harboring related interventions was assessed.

- What are the characteristics of animal studies for interventions known to improve outcomes in humans? Twenty-five interventions shared across humans and animals were identified (supplementary material Table S4). Two further criteria were applied to these 25 interventions. First, multiple studies had to be available in animals and humans. Second, only those interventions in which outcomes were predominantly improved in human studies were selected. Only three interventions survived these criteria – two approved drugs (donepezil and memantine) [58] and exercise. In all three, two characteristics of the cognate animal studies were noted: outcomes were improved in the majority of animal studies and improved outcomes in animals were distributed across a range of different animal models (three or four distinct models or species). The latter finding is reminiscent of recommendations to synthesize data from multiple animal studies to assess likelihood of translation [10], and is the second requirement of the FDA Animal Rule [59]. In the USA, the

Animal Rule is a regulatory means to approve interventions in which efficacy trials in humans are not possible – as such, it provides more-comprehensive guidance regarding efficacy in animals than the traditional Investigational New Drug (IND) exemption [60]. The majority of interventions in animals did not have cognate studies in humans but could still have potential translational significance. To evaluate the repeatability of outcomes across these interventions in animals, the next step was to determine which interventions had been performed across a range of animal models, and what their outcomes were.

- Relatively few interventions have been repeated using different animal models. Eighteen interventions that had been repeated across three or more distinct animal models (supplementary material Table S5) were identified. For each, the network delivered the spectrum of outcomes. Most interventions were associated with a range of outcomes. Further evaluation of the methods of constituent papers was needed to clarify the reasons for this in each case. An example of how intertwined variables rooted in methodology and the species barrier can be seen in studies describing the use of the RXR agonist bexarotene in mice [61–64]. Bexarotene was initially reported to improve cognition and amyloid-beta plaque deposition in three mouse models [61]. Subsequent reports [62–64] were unable to replicate these findings. Potential explanations advanced for this discrepancy included variations in drug formulation [64] and sample sizes that were too small with respect to variability in plaque deposition by age and gender in the model used [62,63,65]. Two interventions in the ‘worsened’ category (streptozotocin and hypoxia) were identified (supplementary material Table S5). Streptozotocin-based interventions across four different animal models support the association between metabolic dysfunction and AD [66], and hypoxic events that worsen the AD-like phenotype support the deleterious role of sleep apnea, stroke or prenatal hypoxia in human AD [67]. As expected, two interventions in the ‘improved’ category were for already approved drugs (donepezil and memantine) [58]. Of the remaining five improved interventions, all were based in the amyloid hypothesis (supplementary material Tables S5,S6).
- Assessing repeatability of mechanistically related interventions. The vast majority of interventions had been examined just once, thus precluding comparison of outcomes across models. However, even though most interventions in animals had been tested only once, several interventions could be mapped to a similar mechanism of action (e.g. phosphodiesterase inhibitors; supplementary material Figure S2). Therefore, the repeatability of outcomes across multiple animal models or species for those mechanisms harboring related interventions was assessed. For this analysis, search criteria were defined as those mechanisms in which related interventions had yielded uniformly improved outcomes in three or more animal models. Five such mechanisms were identified. These, and the related human studies and clinical trials, are given (supplementary material Table S6a). Three novel mechanisms (GO:0004112 – cyclic-nucleotide phosphodiesterase activity; GO:0038180 – nerve growth factor signaling pathway; GO:0048863 – stem cell differentiation) in which all animal studies having improved outcomes were

identified. The remaining two were centered in the amyloid hypothesis, which, together with results of the previous search in which amyloid-based interventions emerged (supplementary material Table S5), prompted a final search to aggregate all animal and human data around the amyloid hypothesis in the dataset. This provided an example of how data regarding a given mechanism could be aggregated in an unbiased manner for subsequent methodology analysis to evaluate pharmacologic and other variables that would impact translatability [23] (supplementary material Table S6b).

Concluding remarks

The analysis described in this paper indicates that the proportion of positive outcomes in animals is far higher than in humans for AD-related interventional studies. Given similar conclusions for other diseases [1,2,21], and the dismal therapeutic reality of AD, this is not a surprise. What is surprising, however, is the magnitude of positive outcome data in animal studies (75% in just one year – 2013). The predictive validity of animal data is the subject of increasing scrutiny [2,3,10,17], and animal models are often criticized for failure to provide translational data [18,68]. The reasons for this reflect enmeshed realities of defining target exposure–response relationships across species [11,12], sources of bias alluded to in the introduction [5,6,17,19–22,69] and the species barrier. In response to identified methodology problems, recent guidelines for improved study design [33,70–73] and reporting [69,74] have emerged, and pressure to apply similar design standards to animal studies [75] and clinical trials is growing. If these reforms are consistently upheld by funding agencies and journal reviewers – this is a work in progress [75] – sources of bias that currently reduce credibility of animal studies will be reduced.

This network-based approach identified strain background as a pervasive potential confounder in mouse studies. Full strain nomenclature was used to define each model in the network. This often required manual searching of methods or references, and background strain was rarely defined. Strain assignment to the identified model using MGI revealed the *rd-1*- and DBA/2-associated blinding alleles, of which the former was noted to confound results of the Morris water maze in a heavily used model of AD [30]. *rd-1* is arguably the best known genetic confounder and information regarding prevalence and detection not only of *rd-1* but for many other confounders in mice has long been available (<http://www.informatics.jax.org/>). Nevertheless, awareness of this allele in our dataset appears to be low. Therefore, it is likely that awareness of the many more subtle confounders and species variations such as those we identified in our dataset [29,34,41,43,44,46] will continue to be low. Solutions to this problem are twofold. First, precise model and strain nomenclature should be required by funding agencies and journal reviewers. Second, although confounding alleles are provided in public databases (<http://www.informatics.jax.org/>), reports describing biological effects of strain-specific confounders on disease mechanisms are buried in the literature. Dissemination of these data would require collation in a comprehensive and user-friendly format that does not rely on prior knowledge to find them.

A second source of variability that, in aggregate, is likely to impact the predictive validity of animal models concerns the relationship between an intervention, engagement of its target, alteration of appropriate biomarkers and altered functional outcome across species.

Outcomes measures across humans and animals are difficult to reconcile for AD. Diagnosis of AD in humans rests primarily on clinical criteria [76]. By contrast, most rodent models are defined by genetic and biomarker-driven criteria (Table 1), and cognitive tests such as the Morris water maze are used to describe functional outcomes [81]. If clinical criteria similar to those defining human AD were used to define mouse models, the clinical progression seen in rodent models at the time of intervention would most probably fall into the category of mild cognitive impairment [77,78]. Therefore, as has been noted for ALS [70], it is likely that interventions in mice occur at a comparatively earlier time point than they do in human clinical trials. Additionally, an intervention could engage its target or change a biomarker, but this does not always correlate with altered functional outcome [79]. Demonstrating this chain of events across multiple species or systems increases confidence for translatability. An example of this is animal studies describing proprotein convertase subtilisin/kexin type 9 (PCKS9) inhibition [80,81]. Whereas observing similar outcomes across models is not itself proof of translatability, it does serve as a proxy marker for those mechanisms or interventions that are robust enough to achieve similar outcomes across different systems. This approach has been proposed by others [10], and is the second prerequisite of the Animal Rule [82]. In this review, this principle was combined with a systems biology networks-based method [83].

By combining these concepts, a method to obtain the following translationally useful information was achieved. First, animal and human data around any intervention or mechanism within the dataset could be rapidly compiled. Second, the network could be interrogated to reveal translationally useful patterns regarding animal model choice, outcomes across species, distribution of mechanisms studied across species or the impact of potential genetic confounders on interpretation of animal data. Lastly, by imposing various criteria (in this case, repeatability of outcome across multiple models), the network could be used to identify those mechanisms (within a landscape of many mechanisms) for which sufficient evidence had been collected in animals to allow further analysis for potential translation to humans. In effect, this approach allowed *in silico* reduction of positive outcomes bias present in animal studies. Once groups of studies have been aggregated by search criteria, the methodology of each must still be scrutinized [23]. The many variables that influence translatability (e.g. PK/PD variables, biomarker validity, time of intervention and termination, selection of outcomes measures across species) are described in the body of each paper, and are best evaluated by investigators, not a search tool.

The failure of animal models to provide translational data is a complex issue. If most animal data cannot be translated, the disconcerting conclusion is that many animal studies are potentially wasted. Tools that can integrate data from extant public databases to reveal translationally useful relationships would improve the utility of available animal model data. Such an overview for any given topic could guide design of related animal studies so as to avoid unnecessary duplication, or address aspects that remain unexamined by existing studies. The approach used to perform this review was developed with semi-automation in

mind. All studies falling within given criteria of initial dataset collection were equally represented (i.e. no filter assessing the validity or quality of a study was imposed). Because of this, initial dataset collection was relatively unbiased in comparison with a traditional review. Because outcome, intervention and species or model are usually reported in the abstract of published studies, these represent variables that are accessible and potentially amenable to automated searching. Lastly, data were arranged around a mechanism-based framework using gene ontology, thus allowing the user to relate studies and outcomes to potential therapeutic targets. Semi-automation of this process would allow users to analyze data that would be periodically updated, and cumulative, thus allowing analysis of progressively larger datasets over time. One source estimates that approximately half of the National Institute of Health (NIH) budget is spent on animal research [78]. The animal model paradigm is unlikely to change soon, but methodology advances [18], changing perspectives [10,17] and economic realities [84] suggest that the days of ‘business as usual’ on this front are over.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NCCR grant 1K26RR025392. The author thanks David Vocolla, Michael Krauthamer and Heather Allore for their helpful discussions.

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 Alzheimer’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)

References

1. Scannell JW, et al. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov.* 2012; 11:191–200. [PubMed: 22378269]
2. Silberberg SD. Should clinicians care about preclinical animal research? *Neurology.* 2013; 80:1072–1073. [PubMed: 23509045]
3. van der Worp HB, et al. Can animal models of disease reliably inform human studies? *PLoS Med.* 2010; 7:e1000245. [PubMed: 20361020]
4. Kilkenny C, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One.* 2009; 4:e7824. [PubMed: 19956596]
5. Moja L, et al. Flaws in animal studies exploring statins and impact on meta-analysis. *Eur J Clin Invest.* 2014; 44:597–612. [PubMed: 24665945]
6. Giles J. Animal experiments under fire for poor design. *Nature.* 2006; 444:981. [PubMed: 17183281]
7. Scott S, et al. Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph Lateral Scler.* 2008; 9:4–15. [PubMed: 18273714]
8. Gill A, et al. No benefit from chronic lithium dosing in a sibling-matched, gender balanced, investigator-blinded trial using a standard mouse model of familial ALS. *PLoS One.* 2009; 4:e6489. [PubMed: 19649300]
9. Tesseur I, De Strooper B. When the dust settles: what did we learn from the bexarotene discussion? *Alzheimers Res Ther.* 2013; 5:54. [PubMed: 24229456]
10. Ioannidis JP. Extrapolating from animals to humans. *Sci Transl Med.* 2012; 4:151ps115.
11. Danhof M, et al. Mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modeling in translational drug research. *Trends Pharmacol Sci.* 2008; 29:186–191. [PubMed: 18353445]
12. Melhem M. Translation of central nervous system occupancy from animal models: application of pharmacokinetic/pharmacodynamic modeling. *J Pharmacol Exp Ther.* 2013; 347:2–6. [PubMed: 23857976]
13. Stephens ML, et al. Accelerating the development of 21st-century toxicology: outcome of a Human Toxicology Project Consortium workshop. *Toxicol Sci.* 2012; 125:327–334. [PubMed: 21948868]
14. Krewski D, et al. New directions in toxicity testing. *Annu Rev Public Health.* 2011; 32:161–178. [PubMed: 21219154]
15. Bhatia SN, Ingber DE. Microfluidic organs-on-chips. *Nat Biotechnol.* 2014; 32:760–772. [PubMed: 25093883]
16. Bueters T, et al. The virtue of translational PKPD modeling in drug discovery: selecting the right clinical candidate while sparing animal lives. *Drug Discov Today.* 2013; 18:853–862. [PubMed: 23665277]
17. Pound P, Bracken MB. Is animal research sufficiently evidence based to be a cornerstone of biomedical research? *BMJ.* 2014; 348:g3387. [PubMed: 24879816]
18. Langley GR. Considering a new paradigm for Alzheimer's disease research. *Drug Discov Today.* 2014; 19:1114–1124. [PubMed: 24662035]
19. Tsilidis KK, et al. Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS Biol.* 2013; 11:e1001609. [PubMed: 23874156]
20. Sena ES, et al. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol.* 2010; 8:e1000344. [PubMed: 20361022]
21. Kyzas PA, et al. Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer.* 2007; 43:2559–2579. [PubMed: 17981458]
22. Macleod MR, et al. Good laboratory practice: preventing introduction of bias at the bench. *Stroke.* 2009; 40:e50–52. [PubMed: 18703798]
23. Fay MP, et al. Anthrax vaccine-induced antibodies provide cross-species prediction of survival to aerosol challenge. *Sci Transl Med.* 2012; 4:151ra126.

24. Bousquet J, et al. System medicine approaches for the definition of complex phenotypes in chronic diseases and ageing. From concept to implementation and policies. *Curr Pharm Des.* 2014; 20:5928–5944. [PubMed: 24641234]
25. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford).* 2011; 2011:baq036. [PubMed: 21245076]
26. Pittler SJ, Baehr W. Identification of a nonsense mutation in the rod photoreceptor cGMP phosphodiesterase beta-subunit gene of the rd mouse. *Proc Natl Acad Sci U S A.* 1991; 88:8322–8326. [PubMed: 1656438]
27. Deacon RM, et al. Aged Tg2576 mice are impaired on social memory and open field habituation tests. *Behav Brain Res.* 2009; 197:466–468. [PubMed: 18977397]
28. Holcomb LA, et al. Behavioral changes in transgenic mice expressing both amyloid precursor protein and presenilin-1 mutations: lack of association with amyloid deposits. *Behav Genet.* 1999; 29:177–185. [PubMed: 10547924]
29. King DL, et al. Progressive and gender-dependent cognitive impairment in the APP(SW) transgenic mouse model for Alzheimer's disease. *Behav Brain Res.* 1999; 103:145–162. [PubMed: 10513583]
30. Yassine N, et al. Detecting spatial memory deficits beyond blindness in tg2576 Alzheimer mice. *Neurobiol Aging.* 2013; 34:716–730. [PubMed: 22819136]
31. Koike H, et al. Disc1 is mutated in the 129S6/SvEv strain and modulates working memory in mice. *Proc Natl Acad Sci U S A.* 2006; 103:3693–3697. [PubMed: 16484369]
32. Anderson MG, et al. Mutations in genes encoding melanosomal proteins cause pigmentary glaucoma in DBA/2J mice. *Nat Genet.* 2002; 30:81–85. [PubMed: 11743578]
33. Howells DW, et al. Bringing rigour to translational medicine. *Nat Rev Neurol.* 2014; 10:37–43. [PubMed: 24247324]
34. Lehman EJ, et al. Genetic background regulates beta-amyloid precursor protein processing and beta-amyloid deposition in the mouse. *Hum Mol Genet.* 2003; 12:2949–2956. [PubMed: 14506131]
35. Couch BA, et al. Delayed amyloid plaque deposition and behavioral deficits in outcrossed AbetaPP/PS1 mice. *J Comp Neurol.* 2013; 521:1395–1408. [PubMed: 23047754]
36. Glazner KA, et al. Strain specific differences in memory and neuropathology in a mouse model of Alzheimer's disease. *Life Sci.* 2010; 86:942–950. [PubMed: 20470790]
37. Hsiao K. Strain dependent and invariant features of transgenic mice expressing Alzheimer amyloid precursor proteins. *Prog Brain Res.* 1998; 117:335–341. [PubMed: 9932419]
38. Wahlsten D, et al. Different rankings of inbred mouse strains on the Morris maze and a refined 4-arm water escape task. *Behav Brain Res.* 2005; 165:36–51. [PubMed: 16191444]
39. Heredia L, et al. Individual housing and handling procedures modify anxiety levels of Tg2576 mice assessed in the zero maze test. *Physiol Behav.* 2012; 107:187–191. [PubMed: 22776622]
40. Strand AD, et al. Conservation of regional gene expression in mouse and human brain. *PLoS Genet.* 2007; 3:e59. [PubMed: 17447843]
41. Loerch PM, et al. Evolution of the aging brain transcriptome and synaptic regulation. *PLoS One.* 2008; 3:e3329. [PubMed: 18830410]
42. Oh S, et al. Reciprocal phylogenetic conservation of molecular aging in mouse and human brain. *Neurobiol Aging.* 2011; 32:1331–1335. [PubMed: 19729224]
43. Inukai S, et al. Novel microRNAs differentially expressed during aging in the mouse brain. *PLoS One.* 2012; 7:e40028. [PubMed: 22844398]
44. Callahan MJ, et al. Augmented senile plaque load in aged female beta-amyloid precursor protein-transgenic mice. *Am J Pathol.* 2001; 158:1173–1177. [PubMed: 11238065]
45. Stargardt A, et al. Reduced amyloid-beta degradation in early Alzheimer's disease but not in the APP^{swe}PS1^{dE9} and 3xTg-AD mouse models. *Aging Cell.* 2013; 12:499–507. [PubMed: 23534431]
46. Frost JL, et al. Pyroglutamate-3 amyloid-beta deposition in the brains of humans, non-human primates, canines, and Alzheimer disease-like transgenic mouse models. *Am J Pathol.* 2013; 183:369–381. [PubMed: 23747948]

47. Ramos-Rodriguez JJ, et al. Rapid beta-amyloid deposition and cognitive impairment after cholinergic denervation in APP/PS1 mice. *J Neuropathol Exp Neurol*. 2013; 72:272–285. [PubMed: 23481704]
48. Doody RS, et al. A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *N Engl J Med*. 2013; 369:341–350. [PubMed: 23883379]
49. Tariot PN, et al. Chronic divalproex sodium to attenuate agitation and clinical progression of Alzheimer disease. *Arch Gen Psychiatry*. 2011; 68:853–861. [PubMed: 21810649]
50. Vigen CL, et al. Cognitive effects of atypical antipsychotic medications in patients with Alzheimer's disease: outcomes from CATIE-AD. *Am J Psychiatry*. 2011; 168:831–839. [PubMed: 21572163]
51. Padala KP, et al. The effect of HMG-CoA reductase inhibitors on cognition in patients with Alzheimer's dementia: a prospective withdrawal and rechallenge pilot study. *Am J Geriatr Pharmacother*. 2012; 10:296–302. [PubMed: 22921881]
52. Schwarz S, et al. Pharmacological treatment of dementia. *Curr Opin Psychiatry*. 2012; 25:542–550. [PubMed: 22992546]
53. Ballard C, Corbett A. Agitation and aggression in people with Alzheimer's disease. *Curr Opin Psychiatry*. 2013; 26:252–259. [PubMed: 23528917]
54. Leong C. Antidepressants for depression in patients with dementia: a review of the literature. *Consult Pharm*. 2014; 29:254–263. [PubMed: 24704894]
55. Laursen B, et al. Cholinergic degeneration is associated with increased plaque deposition and cognitive impairment in APPswe/PS1dE9 mice. *Behav Brain Res*. 2013; 240:146–152. [PubMed: 23178660]
56. Dudkin KN, et al. Disorders of learning and memory processes in a monkey model of Alzheimer's disease: the role of the associative area of the cerebral cortex. *Neurosci Behav Physiol*. 2006; 36:789–799. [PubMed: 16964455]
57. Dudkin KN, et al. Impairments in working memory and decision-taking processes in monkeys in a model of Alzheimer's disease. *Neurosci Behav Physiol*. 2005; 35:281–289. [PubMed: 15875490]
58. Rodda J, Carter J. Cholinesterase inhibitors and memantine for symptomatic treatment of dementia. *BMJ*. 2012; 344:e2986. [PubMed: 22550350]
59. Gronvall GK, et al. The FDA animal efficacy rule and biodefense. *Nat Biotechnol*. 2007; 25:1084–1087. [PubMed: 17921984]
60. Novack GD. Translating drugs from animals to humans: do we need to prove efficacy? *Transl Vis Sci Technol*. 2013; 2:1. [PubMed: 24078898]
61. Cramer PE, et al. ApoE-directed therapeutics rapidly clear beta-amyloid and reverse deficits in AD mouse models. *Science*. 2012; 335:1503–1506. [PubMed: 22323736]
62. Price AR, et al. Comment on "ApoE-directed therapeutics rapidly clear beta-amyloid and reverse deficits in AD mouse models". *Science*. 2013; 340:924-d. [PubMed: 23704553]
63. Veeraghavalu K, et al. Comment on "ApoE-directed therapeutics rapidly clear beta-amyloid and reverse deficits in AD mouse models". *Science*. 2013; 340:924-f. [PubMed: 23704555]
64. Tesseur I, et al. Comment on "ApoE-directed therapeutics rapidly clear beta-amyloid and reverse deficits in AD mouse models". *Science*. 2013; 340:924-e. [PubMed: 23704554]
65. Landreth GE, et al. Response to comments on "ApoE-directed therapeutics rapidly clear beta-amyloid and reverse deficits in AD mouse models". *Science*. 2013; 340:924-g. [PubMed: 23704556]
66. de la Monte SM, Tong M. Brain metabolic dysfunction at the core of Alzheimer's disease. *Biochem Pharmacol*. 2014; 88:548–559. [PubMed: 24380887]
67. Daulatzai MA. Death by a thousand cuts in Alzheimer's disease: hypoxia--the prodrome. *Neurotox Res*. 2013; 24:216–243. [PubMed: 23400634]
68. Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. *JAMA*. 2006; 296:1731–1732. [PubMed: 17032985]
69. Kilkenny C, et al. Animal research: reporting in vivo experiments: the ARRIVE guidelines. *Br J Pharmacol*. 2010; 160:1577–1579. [PubMed: 20649561]

70. Kong Q, et al. The importance of preclinical trial timing - a potential reason for the disconnect between mouse studies and human clinical trials in ALS. *CNS Neurosci Ther.* 2012; 18:791–793. [PubMed: 22712693]
71. Henderson VC, et al. Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for *in vivo* animal experiments. *PLoS Med.* 2013; 10:e1001489. [PubMed: 23935460]
72. Perrin S. Preclinical research: make mouse studies work. *Nature.* 2014; 507:423–425. [PubMed: 24678540]
73. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature.* 2012; 483:531–533. [PubMed: 22460880]
74. Landis SC, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature.* 2012; 490:187–191. [PubMed: 23060188]
75. Baker D, et al. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol.* 2014; 12:e1001756. [PubMed: 24409096]
76. McKhann GM, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:263–269. [PubMed: 21514250]
77. Albert MS, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011; 7:270–279. [PubMed: 21514249]
78. Puzzo D, et al. Behavioral assays with mouse models of Alzheimer's disease: practical considerations and guidelines. *Biochem Pharmacol.* 2014; 88:450–467. [PubMed: 24462904]
79. LaClair KD, et al. Treatment with bexarotene, a compound that increases apolipoprotein-E, provides no cognitive benefit in mutant APP/PS1 mice. *Mol Neurodegener.* 2013; 8:18. [PubMed: 23764200]
80. Frank-Kamenetsky M, et al. Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. *Proc Natl Acad Sci U S A.* 2008; 105:11915–11920. [PubMed: 18695239]
81. Chan JC, et al. A proprotein convertase subtilisin/kexin type 9 neutralizing antibody reduces serum cholesterol in mice and nonhuman primates. *Proc Natl Acad Sci U S A.* 2009; 106:9820–9825. [PubMed: 19443683]
82. Snoy PJ. Establishing efficacy of human products using animals: the US food and drug administration's "animal rule". *Vet Pathol.* 2010; 47:774–778. [PubMed: 20551476]
83. Vidal M, et al. Interactome networks and human disease. *Cell.* 2011; 144:986–998. [PubMed: 21414488]
84. Macilwain C. Funding in 2012: "great recession" starts to bite. *Cell.* 2012; 148:14–16. [PubMed: 22265395]

Highlights

- 448 interventions across 752 studies for Alzheimer's disease were compared
- A networks-based systems biology approach was used
- Repeatability of outcomes across species by intervention and mechanism was assessed
- 75% of animal studies reported an improved outcome
- Strain background was a significant potential confounder
- 5% of interventions had been tested across species or across multiple models
- This approach supports *in silico* reduction of positive outcomes bias in animal studies

Figure 1.

Interventional outcomes across humans and animal studies. **(a)** Combined human and animal network. The majority of human clinical trials have no available results (yellow nodes). The majority of animal studies (blue nodes, 75.9%) reported an improved outcome. **(b,c)** Individual outcomes networks for humans (b) or animals (c). Just under half of human studies (b, blue nodes; 46.6%) over a five year period report improvement. Animal studies with a worsened outcome (c, red nodes; 13.6%) were predominantly studies where the intent was to worsen some aspect of the AD-like phenotype. Therefore, these are not directly comparable with the few human studies in which an intervention intended to treat the disease actually worsened outcome (b, red nodes; 2.7%). Interventions producing no effect constitute over a quarter of human studies over five years (b, orange nodes; 26.4%) but only a small percentage of animal studies report no effect (c, orange nodes; 4.8%).

Figure 2.

Mechanisms of disease in AD across humans and animals. The first 16 mechanisms shared across humans and animals (supplementary material Table S5) are compared using a stacked 100% bar. Studies targeting neurotransmitter systems are indicated by a thin horizontal black bar and constitute the first eight mechanisms in Table S5. These predominate in humans compared with animals. Within this category, studies targeting the cholinergic system (bright orange; GO:0015464 – acetylcholine receptor activity) predominate in humans and animals. Studies focused on beta-amyloid-related interventions were similarly distributed across species (blue shades, thick horizontal grey bar). Interventions in which outcomes describe inflammation (gray), oxidative stress (red) and glucose homeostasis (purple) prevail in animals. Full data are given in Table S5.

Table 1

Most heavily used animal models in the network

Animal model	Studies (%)	Human allele(s) ^a	Strain background ^a
APP/PS1: Tg(APPswe, PSEN1dE9) 85Dbo. PMID: 11337275	85 (24%)	APP KM595/596NL Swedish PSEN1 deltaE9	(C57BL/6 x C3H)F2
APP695swe: Tg(APPswe)2576Kha. PMID: 8810256	42 (12%)	APP K670N/M671L Swedish	(C57BL/6 x SJL)F1
3xTg-AD: B6.Cg-Psen1tm1Mpm Tg(APPswe, tauP301L)1Lfa. PMID: 12895417	38 (11%)	APP K670N/M671L Swedish PSEN1 MAPT P301L	C57BL/6J
5XFAD: Tg(APPswF1Lon, PSEN1* M14 6L* L286V)6799Vas. PMID: 17021169	19 (5%)	APP K670N/M671L Swedish APP V717I London APP I716V Florida PSEN1 M146L PSEN1 L286V	(C57BL/6 x SJL)F1
Rodent: cerebral injection of amyloid beta oligomers. PMID: 15810904	17 (5%)	None	Inbred/outbred strains
APPPS1-21: Tg(Thy1-APPsw, Thy1-PSEN1* L166P)21Jckr. PMID: 16906128	14 (4%)	APP K670N/M671L Swedish PSEN1 L166P	C57BL/6J
CRND8: Tg(PRNP-APPsweInd)8Dwt. PMID: 11279122	14 (4%)	APP K670N/M671L Swedish APP V717F Indiana	(C3H/HeJ x C57BL/6J)F1
PDAPP-J20: Tg(PDGFB-APPswInd)20Lms. PMID: 10818140	13 (4%)	APP K670N/M671L Swedish APP V717F Indiana	(C57BL/6 x DBA/2)F2
TASTPM: Tg(APPswe)2576Kha/Tg(PDGFB-PSEN1M146L). PMID: 15261108	9 (3%)	APP K670N/M671L Swedish PSEN1 M146L	(C57BL/6 x SJL)F1SW x (C57BL/6 x DBA/2)F1
APP23: Tg(Thy1-APP)3Somm. PMID: 9371838	8 (2%)	APP K670N/M671L Swedish	(C57BL/6J x DBA/2)F1
SAMP8 mice. PMID: 24269312	8 (2%)	None	AKR/J
PDAPP: Tg(APPV717F)109Ili. PMID: 7845465	6 (2%)	APP V717F Indiana	Not specified
APP/PS1: Tg(Tg(Thy1-APPLon)2Vln, Thy1-PSEN1* A246E)2Vln. PMID: 10964951	5 (1%)	APP V717I London PSEN1 A246E	FVB/N
APP751: Tg(Thy1-APPswDutLon)101Lpr. PMID: 14637096	5 (1%)	APP KM595/596NL Swedish APP E618Q Dutch APP V6421 London	CBA x C57BL/6
Tg(Thy1-APPswDutIowa)Bwevn. PMID: 14985348	5 (1%)	APP E693Q/D694N vasculotropic Dutch/Iowa APP K670N/M671L Swedish	C57BL/6
APP/PS1: B6C3H-Tg (APP695) 3Dbo Tg (PSEN1 A246E) 5Dbo/J; PMID: 9354339	4 (1%)	APP KM595/596NL Swedish PSEN1 A246E	(C57BL/6J x C3H/HeJ)F2

Each model is defined by three identifiers: a synonym, the precise strain nomenclature and the PMID of the seminal manuscript describing it.

^aData from Mouse Genome Informatics (MGI; <http://www.informatics.jax.org>)

Strains carrying the *rd-1* mutation (red): ABJ/LeJ; BDP/J; BUB/BinJ; C3H and all substrains; CBA/J; CBA/N; FVB/NJ; JGBF/LeJ; MOLD/RkJ; MOLF/EiJ; NFS/N; NON/LtJ; P/J; PL/J; RSV/LeJ; SB/LeJ; SF/CamEi; SF/CamRk; SK/CamEi; ST/bJ; SJL/J; SWR/J; WB/ReJ; WC/ReJ (<http://eyemutant.jax.org/rd1.html>).

Table 2

Outcomes of interventional studies for Alzheimer's disease (AD) across species

	Animal studies (2013)	Human studies (2013)	Human studies (2009–2013)
Improved	268 (75.9%)	24 (57.1%)	69 (46.6%)
Mixed outcome	22 (6.2%)	4 (9.5%)	13 (8.8%)
No effect	14 (3.9%)	7 (16.6%)	39 (26.4%)
No adverse effect	0	6 (14.3%)	23 (15.5%)
Worsened	49 (13.8%)	1 (2.4%)	4 (2.7%)
Total	353 (100%)	42 (100%)	148 (100%)