Full length article

Are some animal models more equal than others? A case study on the translational value of animal models of efficacy for Alzheimer's disease

Désirée H. Veening-Griffioen[a],*, Guilherme S. Ferreira[a], Peter J.K. van Meer[a,b], Wouter P.C. Boon[c], Christine C. Gispen-de Wied[b], Ellen H.M. Moors[c], Huub Schellekens[a]

[a] Utrecht Institute of Pharmaceutical Sciences, Department of Pharmaceutics, Utrecht University, Universiteitsweg 99, 3584 CG, Utrecht, the Netherlands
[b] Medicines Evaluation Board, Graadt van Roggenweg 500, 3531 AH, Utrecht, the Netherlands
[c] Copernicus Institute of Sustainable Development, Innovation Studies, Utrecht University, Princetonlaan 8a, 3584 CB, Utrecht, the Netherlands

A B S T R A C T

Clinical trial failures (> 99%) in Alzheimer's disease are in stark contrast to positive efficacy data in animals. We evaluated the correlation between animal and clinical efficacy outcomes (cognition) in Alzheimer's disease using data from registered drugs as well as interventions tested in phase II or III clinical trials for Alzheimer's disease.

We identified 20 interventions, which were tested in 208 animal studies in 63 different animal models. Clinical outcome was correlated with animal results in 58% of cases. But, individual animal models showed divergent results across interventions, individual interventions showed divergent results across animal models, and animal model outcomes were determined with 16 different methods.

This result is unsurprising due to poor external validity (what do we model) of the animal models. Although the animal models all share Alzheimer's disease symptoms, none represents the whole syndrome. Investigators did not motivate why one model was chosen over another, and did not consider the ways the disease phenomena were generated (spontaneous, (experimentally) induced or by genetic modification), or the species characteristics, which determine the outcomes. The explanation for the lack of correlation between animal and human outcomes can be manifold: the pathogenesis of Alzheimer's disease is not reflected in the animal model or the outcomes are not comparable.

Our conclusion is that currently no animal models exist which are predictive for the efficacy of interventions for Alzheimer's disease.

1. Introduction

With a failure rate of more than 99%, drug development in Alzheimer's disease has been largely unsuccessful: since 2002 over than 400 clinical trials failed due to lack of efficacy (Cummings et al., 2014; Rinaldi, 2018). One of the reasons for these failures was because data obtained from animal models did not translate to the clinic (Anand et al., 2017). This situation has led to the question of whether animal models for Alzheimer's disease are useful at all, especially in evaluating the efficacy of new drugs.

The poor translation of animal model data has been attributed mainly to low internal validity (van Meer et al., 2015). Animal studies are frequently poorly designed (e.g. non-randomized, non-blinded), minimally standardized and underpowered (Egan et al., 2016; Howells et al., 2014; Ioannidis et al., 2014; van der Worp et al., 2010). In addition, the design of animal studies is often inadequately described, and positive outcomes are selectively reported (Begley and Ioannidis, 2015; Fanelli, 2012), making the results unreliable. Internal validity can easily be improved with better designing and reporting, for instance, by compliance with guidelines such as the Animal Research Reporting of *In Vivo* Experiments (ARRIVE) (Kilkenny et al., 2010) and the Planning Research and Experimental Procedures on Animals: Recommendations for Excellence (PREPARE) (Smith et al., 2018). Preregistration of animal studies (e.g. preclinicaltrial.eu) has been suggested to lead to more transparency on study design, reducing duplication of studies and risk of selective outcome reporting (Ritskes-Hoitinga and Wever, 2018).

In contrast with the interest in internal validity, the external validity of animal models, i.e. how the results obtained in the model can be generalized to the human disease, is barely receiving attention (van Meer et al., 2015). And even for marketing authorization applications, investigators are not required to substantiate the relevance of the animal models used significantly, nor are there systematic and

standardized methods to do so.

Animal models of efficacy should simulate the disease characteristics, including symptoms, response to treatment and similar diagnostic outcome assessment tools (Sams-Dodd, 2006). These animal models can be categorized in the way the disease characteristics are acquired: spontaneous (by nature or natural occurring genetic variants), induced (chemical, biological or physical) or by genetic modification (transgenic) (Hau, 2008).

Animal models can be compared to the human disease according to different aspects: face validity (the similarity in clinical symptoms), construct validity (the similarity in etiology) and predictive validity (the similarity in the clinical response to treatment) (McKinney and Bunney, 1969; Sams-Dodd, 2006; Willner, 1984).

For Alzheimer's disease, especially the construct validity cannot be fully defined because the etiology of the disease is still unclear. The four main hypotheses (aging, neurotransmitter, amyloid pathway, and neuroinflammation), which involve multiple targets, have yet to be confirmed. In this study, we investigated the correlation between human and animal outcomes of interventions for Alzheimer's disease, to assess whether some animal models of efficacy are more likely to correlate with human outcomes than others.

2. Material and methods

2.1. Identification of clinical trials for Alzheimer's disease

Phase III trial data were included from the drugs registered in the European Union and the United States of America for Alzheimer's disease. These data are described in the summary review of scientific data used for marketing authorization, which are available on the websites of both the European Medicines Agency (EMA) and the Food and Drug Administration (FDA). Data were included for four Alzheimer's disease drugs on the market, i.e.: donepezil (United States Food and Drug Administration, 2006), galantamine (United States Food and Drug Administration, 2001), memantine (European Medicines Agency, 2004) and rivastigmine (European Medicines Agency, 2005).

A targeted search was performed on www.clinicaltrials.gov to identify all phase III clinical trials intended to assess the efficacy of a chemical, biological or nutritional intervention in Alzheimer's disease based on the following search criteria: Alzheimer (condition or disease), phase III, interventional, monotherapy arm, randomized, double-blind and placebo-controlled. Trials were included that had published results (including statistical significance testing) on cognitive outcome measures, which is one of the accepted parameters to show efficacy of an intervention for Alzheimer's disease (European Medicines Agency, 2018). We chose cognition as the parameter to compare human and animal efficacy data, as cognitive outcome can be objectively tested in both clinic and in animal studies and there are far fewer correlates for other (subjective) outcome measures.

For trials without reported results in clinicaltrials.gov database, we searched on PubMed on www.ncbi.nlm.nih.gov/pubmed for publications of trial data using the Trial ID: "NCTxxxxxxxx"[All Fields]. Publications were included up to 28 May 2018. For all trials, we extracted data for the parameters listed in Table 1.

2.2. Identification of animal models for Alzheimer's disease

Animal studies for Alzheimer's disease were identified in PubMed using the method from Leenaars and colleagues (Leenaars et al., 2012). This method systematically identifies all relevant animal studies. The following search components (SCs) were obtained and combined: SC1 intervention (including synonyms as published on PubChem on www.pubchem.ncbi.nlm.nih.gov), SC2 disease of interest (Alzheimer's disease as reported in the PubMed Medical Subject Headings (MeSH)-tree), SC3 animal species by the PubMed animal filter as published by Hooijmans and colleagues (Hooijmans et al., 2010) and SC4 cognition

(and synonyms). A detailed overview of the search components is listed in S1 Supplementary data.

Animal studies were included if a cognitive outcome was measured, the study was placebo-controlled, the study had a monotherapy arm of one of the interventions identified in Section 2.1, and the intervention started after induction of cognitive impairment. Publications were included from inception up to 28 May 2018.

Animal models were grouped according to Hau's classification (Hau, 2008), based on how the cognitive impairment was obtained, which could be either spontaneous (mutation), induced (chemical, biological or physical) or by genetic modification (transgenic).

For all animal studies, we extracted data from the parameters listed in Table 1.

2.3. Reporting quality

We evaluated the compliance to an adapted version of the ARRIVE guidelines (Kilkenny et al., 2010), as an indication of the internal validity of the included animal studies.

We evaluated reporting on the following 20 parameters: population characteristics (species, strain, age and sex), housing characteristics (type of facility, type of cage, type of bedding and number of cage companions), husbandry characteristics (light/dark cycle, temperature/humidity, quality of water, type of food, access to food/water, duration of acclimatization), and design characteristics (sample size, sample size calculation, randomization to cage and randomization to treatment, allocation concealment and blinded outcome assessment). These parameters were assessed as reported (Y) or not reported (N). Compliance with ARRIVE was calculated using the formula:

$$Compliance\ (\%) = \frac{number\ of\ reported\ parameters}{20} \times 100$$

2.4. Analysis

For the data we extracted in Section 2.1. and 2.2. we calculated the overall outcome of individual trials or animal studies (Section 2.4.1.), the overall outcome of all trials or all animal studies of a specific animal model (Section 2.4.2.). Additionally, for each individual intervention (Section 2.4.3.) and across interventions (Section 2.4.4.), we compared clinical outcome to the outcome in the animal model.

2.4.1. Overall clinical trial or animal study outcome for an intervention

Characteristics of cognitive impairment in Alzheimer's disease include attention, working memory, executive function, episodic memory, semantic memory, visual and verbal memory, processing speed, procedural memory, social cognition and language (Millan et al., 2012). These characteristics can be assessed by different methods (both in humans and in animals), and for each characteristic different parameters can be reported. We treated these parameters equally, since to our knowledge, there is insufficient data ranking or comparing these parameters within and between species.

We used outcome results significance testing as reported by the study authors. We reported an overall cognitive outcome of an individual animal study. The overall study cognitive outcome ratio was calculated using the formula:

$$Overall\ study\ cognitive\ outcome\ (ratio)$$
$$= \frac{number\ of\ 'improved'cognitive\ parameters}{total\ number\ of\ cognitive\ parameters}$$

The overall animal study outcome can be either "improved" (when ratio > 0.5), "mixed" (when ratio = 0.5) or "no effect" (when ratio is < 0.5).

Table 1
Parameters for data extraction from clinical trials and animal studies.

Parameter	Description
Intervention	Drug or food component tested in trial or study
PubChemID	Unique identifier for an intervention as listed in the PubChem compound database at www.pubchem.ncbi,nlm.nih.gov
ATC index (4th level)	Drug classification as listed in the WHO ATC/DDD index at www.whocc.no/atc_ddd_index
Pharmacological target	Pharmacological target as listed in the PubChem compound database: anti-cholinergic/anti-glutaminergic, amyloid/tau/secretase pathway, anti-neuroinflammation or other
Market authorization	Market authorization status in the European Union or in the United States of America: approved (Y) or not approved (N)
Market authorization (year)	Year of first market approval
Trial or Model Classification	Classification of Trial or Model - phase III or II (clinical trial) - spontaneous (SP), induced (IN) or genetically modified (GM) (animal studies)
Trial or Model ID	Unique name for trial or animal model; an abbreviation composed of classification, species and trial/model
Publication (year)	Year the trial or study results were published
Trial or Study ID	Unique trial or study identifier by - Trial code or Trial NCT-number (clinical trials) - Ascension number of publication in PubMed (animal studies) at www.ncbi.nlm.nih.gov
Trial or Model Name	Name for trial or animal model. Animal models are further described in Table 4
Trial of Model Motivation	Reason for trial or choice of model: efficacy (clinical) and previous research, model of cognitive impairment, model of aging-induced cognitive impairment, model of cognitive impairment or other hallmarks of Alzheimer's disease (animal model)
Population - species	Species used in trial or study grouped by their scientific name
Population - strain	Strain disclosure of study object
Population - strain disclosure	Authors' reporting of strain: reported (Y) or not reported (N)
Population - sex	Sex disclosure of study object; male, female, all (both sexes) or not reported
Population - sex disclosure	Authors' reporting of sex: reported (Y) or not reported (N)
Population - age	Age disclosure of study object
Population - age (unit)	Unit in which age of the study object was measured
Population - age disclosure	Authors' reporting of age: reported (Y) or not reported (N)
Outcome Classification	Outcome classification in trial or study: cognitive, functional, behavior and mood, quality of life, global, economical, pathological score, biomarker – safety or biomarker – pharmacodynamic response
Outcome Result	Trial or study outcome result according to the study author's significance test: worsened, no effect or improved compared to control
Outcome Measure	Method used to determine the outcome
Outcome Description	Method parameter used to measure the outcome, and when reported by the study's author whether the outcome was: - primary; the outcome in a trial or study which has the greatest or most desirable therapeutic effect, this parameter was used to calculate the sample size - secondary; the outcomes are used to evaluate additional effects of the intervention not included in the primary outcome measure - exploratory; the outcomes are used to evaluate potential new outcomes for a specific indication
Comments	Any other information

2.4.2. Overall clinical or animal model outcome for an intervention

Interventions can be tested in several clinical trials or in different studies with the same animal model. We reported the overall cognitive outcome of an intervention in a trial or animal model. We used outcome results significance testing as reported by the study authors. The overall model cognitive outcome ratio for an intervention was calculated using the formula:

$$\text{Overall model cognitive outcome (ratio)} = \frac{number\ of\ 'improved'\ studies}{total\ number\ of\ studies}$$

The overall outcome of an intervention can be either "improved" (when ratio > 0.5), "mixed" (when ratio = 0.5) or "no effect" (when ratio is < 0.5).

2.4.3. Predictive value of a specific intervention in a specific animal model for clinical outcome

For each intervention, we compared the overall outcome in an animal model to the clinical outcome. When the overall outcome was the same in both the clinic and animal model, the animal model's prediction of the clinical outcome was considered "correct", and if not, as "incorrect". We reported "unclear" when the overall model outcome (section 2.4.2.) was 'mixed'.

2.4.4. Predictive value of an animal model across interventions for clinical outcome

For all interventions tested in each animal model, we compared their overall outcome to their clinical outcome. The overall model prediction of the clinical outcome (across interventions) was calculated using the formula:

$$\text{Overall model prediction for clinical outcome (\%)}$$
$$= \frac{number\ of\ 'correct'\ (or\ 'unclear'or\ 'incorrect')\ interventions}{total\ number\ of\ interventions} x\ 100$$

3. Results

3.1. Clinical trials for Alzheimer's disease

We identified 8 phase III trials for the registered drugs, published in the EMA or FDA summary review of scientific data used for marketing authorization. A total of 134 completed interventional phase III clinical trials for Alzheimer's disease were registered on clinicaltrials.gov. Of these, 37 (28%) reported results in the database, while 97 (72%) did not. For the trials without results, the PubMed search identified another 38 publications. In total 30 phase III trials, in which 20 different interventions were tested, met the inclusion criteria (Fig. 1A).

Via the same process, we identified 225 phase II trials on clinicaltrials.gov. Of these, 60 (27%) reported results in the database, while 165 (73%) did not. For the trials without results, the PubMed search identified another 42 publications. However, we only included the trials in which the interventions identified in phase III were tested. In total 6 phase II trials, in which 4 of the 20 different interventions were tested, met the inclusion criteria (Fig. 1B).

For the 36 clinical trials included in our study, 18 trials were included for which no data were published in clinicaltrials.gov but found in PubMed. For these trials the PubMed ID (PMID) from where the data were extracted were added in S3 Supplementary data (Comments; Column AQ). Additionally we checked for anticipated discrepancies of primary and secondary outcome measures between the clinicaltrials.
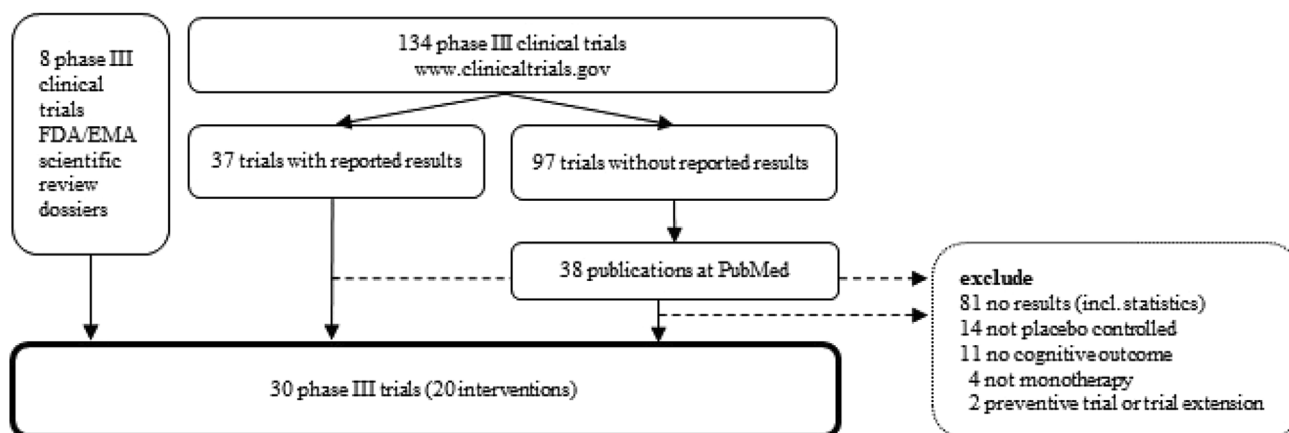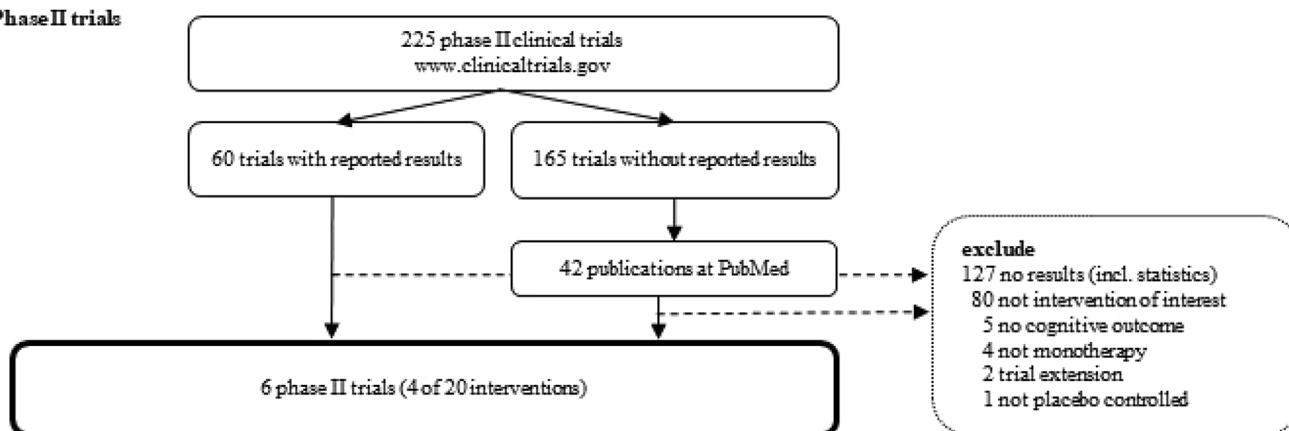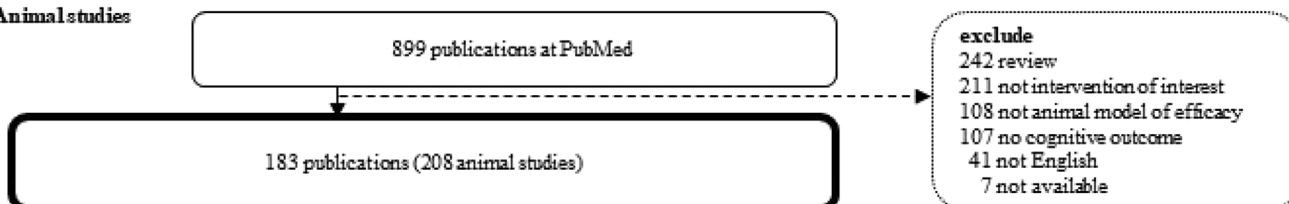
A. Phase III trials

B. Phase II trials

C. Animal studies

Fig. 1. Flow diagram of the selection of phase III trials (A), phase II trials (B) and animal studies (C).

gov database and PubMed (Hartung et al., 2014).

Of these 18 trials, 15 trials reported the same primary or secondary outcome measures in PubMed and in clinicaltrials.gov (Table 2). For 3 trials the information on primary or secondary outcome measures was only reported in PubMed. An overview of the included phase II and III trials, including information on methods used to assess cognition and the outcome, is shown in Table 2. Cognitive outcomes were generally assessed with one of the Alzheimer's Disease Assessment Scale, cognitive subscale (ADAS-cog subscales) the Minimal Mental State Examination (MMSE) or the Severe Impairment Battery (SIB). Improved clinical cognitive outcomes were reported for interventions targeting the cholinergic/glutaminergic pathway, which included all the currently registered drugs and the amyloid/tau/secretase pathway.

The overall cognitive outcome per intervention as well as the number of trials and the market authorization status in 2018 are listed in Table 3. All interventions showing overall improved cognitive outcome received market authorization for Alzheimer's disease. Raw data concerning the trials is available in S3 Supplementary data.

3.2. Animal models of efficacy for Alzheimer's disease

A schematic representation of the targeted search for animal studies is shown in Fig. 1C. We identified 899 publications of which 668 (74%) did not meet the inclusion criteria, 41 (4%) were not in the English language, and 7(1%) were not available via the University Library. Some of the 183 included publications, published between 1988 and 2018, reported more than 1 animal study, e.g. 1 intervention tested in different animal models. The data from 208 individual animal studies were extracted (S3 Supplementary data).

Of the 20 identified interventions (Table 3), 13 were tested in 1 or more of the 208 animal studies. The animal studies were conducted in 7 different species: 1 (1%) was in *Canis lupus familiaris* (dog), 2 (1%) in *Drosophila melanogaster* (fruit fly), 2 (1%) in *Microcebus murinus* (lemur), 5 (2%) in *Oryctolagus cuniculus* (rabbit), 7 (3%) in *Macaca mulatta* (rhesus monkey), 63 (30%) in *Rattus norvegicus* (rat) and 127 (61%) in *Mus musculus* (mouse).

Table 2

Overview of interventional clinical trials with reported cognitive outcome measures for Alzheimer's disease. Objective tests for cognition used in clinical trials are (Dekker et al., 2019): Animals Category Fluency Assessment (ACF), Alzheimer's Disease Assessment Scale - cognitive subscale (ADAS-cog), Clock Drawing Assessment (CDT), Digit Symbol Substitution Test (DSST), Verbal Fluency Assessment (FAS), Minimal Mental State Examination (MMSE), One Card Learning Test (OCL), Severe Impairment Battery (SIB), Trail Making Test (TMT) and Wechsler Adult Intelligence Scale (WIAS). Data show the outcome of the trials: primary (P) or secondary (S).

Intervention classification	Intervention(s)	Clinical trial (phase; year)	Cognitive outcome	Cognitive outcome no effect (or * worsened) compared to control	Trial ID
Cholinergic/glutaminergic	donepezil	II; 2011		ADAS-cog (P)	NCT00348192
		II; 2014	ADAS-cog (P), MMSE (S)		NCT01137526
		II; 2015		ADAS-cog (P)	NCT00948909
		II; 2017	OCL (P)		NCT02064920
		III; 2006	ADAS-cog (P), SIB (P)		NDA 20–690/S-026
		III; 2017	MMSE (S)	ADAS-cog (P)	NCT00428090
	galantamine	III; 2000	ADAS-cog (P)		FDA:21-168
		III; 2009	SIB (S)		NCT00216593
		III; 2012	ADAS-cog (P)		NCT00814801
	memantine	III; 1994	ADAS-cog (P)	MMSE (S)	MRZ-9202
		III; 1996	ADAS-cog (P), MMSE (S)		MRZ-9408
		III; 1999	SIB (S)	MMSE (S)	MRZ-9605
		III; 2010	SIB (S)		NCT00322153
		III; 2014		ADAS-cog (P), MMSE (P)	NCT00235716
	rivastigmine	III; 2005	ADAS-cog (P), MMSE (S)		B303
		III; 2005	ADAS-cog (P), MMSE (S)		B351
		III; 2005	ADAS-cog (P), MMSE (S)		B352
		III; 2007		ADAS-cog (P)	NCT00000174
		III; 2007	ADAS-cog (P)		NCT00099242
Amyloid/Tau/secretase	bapineuzumab	III; 2014		ADAS-cog (P), MMSE (S)	NCT00574132
		III; 2014		ADAS-cog (P), MMSE (S)	NCT00575055
	IVIG	III; 2014		ADAS-cog (S), MMSE (S), WIAS (S), FAS (S), ACF (S), TMT (S), CDT (S)	NCT00818662
	LMTM	III; 2016		MMSE (S)	NCT01689246
	semagacestat	II; 2008		SIB (S)	NCT00244322
		III; 2014	ADAS-cog (P), MMSE (S)		NCT00594568
		III; 2014		ADAS-cog (P), MMSE (S)	NCT00762411
	solanezumab	III; 2014	MMSE (S)	ADAS-cog (S)	NCT00904683
		III; 2014		ADAS-cog (P), MMSE (S)	NCT00905372
Neuroinflammation	celecoxib	III; 2008		MMSE (S)*	NCT00007189
	folate	III; 2008		ADAS-cog (P), MMSE (S)	NCT00056225
	indomethacin	III; 2008		ADAS-cog (P), MMSE (S)	NCT00432081
	naproxen	III; 2008		MMSE (S)*	NCT00007189
	rosiglitazone	II; 2010		SIB (S)	NCT00265148
		III; 2017		ADAS-cog (P), MMSE (S), DSTT (S)	NCT00348140
		III; 2017		ADAS-cog (P), MMSE (S)	NCT00428090
	tarenflurbil	III; 2009		ADAS-cog (P)	NCT00105547
	vitamin B6	III; 2008		ADAS-cog (P), MMSE (S)	NCT00056225
	vitamin B12	III; 2008		ADAS-cog (P), MMSE (S)	NCT00056225
	vitamin E	III; 2014		ADAS-cog (P), MMSE (P)	NCT00235716
Anxiety/depression	aripiprazole	III; 2012	MMSE (S)	MMSE (S)	NCT01438060
	trazodone	III; 2014		MMSE (S), DSTT (S)	NCT01142258

3.3. Variation in animal models of efficacy for Alzheimer's disease

The 208 animal studies were performed in 63 different animal models (Table 4). These animal models were classified by the way the cognitive impairment was acquired. We found 9 spontaneous models (natural cognitive impairment), 36 induced models (chemical, biological or physical induced cognitive impairment) and 18 genetically modified models (cognitive impairment induced by genetic modification). Of these animal models, 33 were used only once, and therefore excluded in the further analysis.

The use of the different model classes changed over time (Fig. 2.). Spontaneous models were used in comparable numbers over the years, while the use of both induced models and genetically modified models increased exponentially after 2000.

Since multiple animal models were used to evaluate the efficacy of interventions for Alzheimer's disease, we investigated the authors' motivation for their animal model choice (Fig. 3.). For every animal study the authors motivated their choice of animal model, which is listed in S3 supplementary data. The author's motivations were then classified into the following categories: the model shows aging-related cognitive impairment (15%), the model shows cognitive impairment as well as other markers of Alzheimer's disease (17%), the model shows cognitive impairment (33%), or the model was previously used (35%).

3.4. Unclear prediction of clinical outcome by animal models of efficacy for Alzheimer's disease

We compared the overall outcome of each intervention-model combination to the clinical outcome. We found 74 intervention-model combinations. Individual data is available in S4 Supplementary data. The percentage of animal model's prediction of clinical outcomes is shown in Fig. 4A. In general, 58% of animal models predicted correctly for clinical outcome of an intervention. Per animal model class, 23% of spontaneous models, 73% of induced models, and 58% of genetically

Table 3
Overall clinical trial outcome (improved, mixed or no effect) for 20 intervention as well as their market authorization status in 2018. Data show number of trials and the ratio of improved trials per total amount of trials.

Intervention classification	Intervention	Phase II/III trials (number)	Overall phase II/III Trial outcome (ratio; result)		Market authorization (Y/N)
Cholinergic/glutaminergic	donepezil	6	0.5;	mixed	Y
	galantamine	3	1.0;	improved	Y
	memantine	5	0.6;	improved	Y
	rivastigmine	5	0.8;	improved	Y
Amyloid/Tau/secretase	bapineuzumab	2	0.0;	no effect	N
	IVIG	1	0.0;	no effect	N
	LMTM	1	0.0;	no effect	N
	semagacestat	3	0.3;	no effect	N
	solanezumab	2	0.5;	mixed	N
Neuroinflammation	celecoxib	1	0.0;	no effect	N
	indomethacin	1	0.0;	no effect	N
	folate	1	0.0;	no effect	N
	naproxen	1	0.0;	no effect	N
	rosiglitazone	3	0.0;	no effect	N
	tarenflurbil	1	0.0;	no effect	N
	vitamin B6	1	0.0;	no effect	N
	vitamin B12	1	0.0;	no effect	N
	vitamin E	1	0.0;	no effect	N
Anxiety/depression	aripiprazole	1	0.5;	mixed	N
	trazodone	1	0.0;	no effect	N

modified models correlated with the clinical outcome of an intervention.

The best correlations with clinical outcome were obtained in animal models in which interventions of the cholinergic/glutaminergic class (the currently registered drugs) were tested (Fig. 4B.). For these type of interventions 72% of the animal studies predicted the clinical outcome correctly, while 19% predicted the clinical outcome incorrectly (Fig. 5.). The animal studies, in which interventions of the cholinergic/ glutaminergic class were tested, and which predicted the clinical outcome incorrectly, were often performed in animals models based on the amyloid pathway hypothesis or aging-related hypotheses.

The number of studies and the number of interventions carried out in each animal model is shown in Fig. 6. The overall outcome of each animal model is compared to the clinical outcomes across interventions (Fig. 6). While 10 animal models correctly predicted the clinical outcome of all tested interventions, in 2 animal models not a single prediction proved correct. A total of 18 animal models partly predicted the clinical outcome.

The number of animal models tested with each intervention is shown in Fig. 7. The overall outcome of each intervention is compared to the clinical outcomes across animal models. While the outcome of 2 interventions correctly predicted the clinical outcome, 3 interventions had not a single correct prediction. A total of 5 interventions showed a mixed pattern.

3.5. Variations in outcome assessment methodology

The cognitive outcome was assessed by 16 different methods (Fig. 8.). The most often used methods were: Morris Water Maze (37%), the Passive Avoidance Test (27%), Object Recognition Task (14%) and the T- or Y-maze Alteration Test (15%). Detailed information on these methods is available in S2 Supplementary data.

Besides cognitive parameters, functional and behavioral parameters were measured in 10 (5%) animal studies.

3.6. Quality of animal studies between 1988 and 2017

The compliance to the 20 aspects of the ARRIVE guidelines was evaluated (Table 5). Population characteristics and sample size were most often disclosed (> 88%), while many aspects of housing and husbandry, blinding and randomization were not reported (< 24%). Furthermore, between 1988 and 2018 the overall compliance level

increased minimally (Fig. 9.).

4. Discussion

Animal models of efficacy for Alzheimer's disease correlated in 58% of cases with the clinical outcomes (Fig. 4A.). However, the value of these animal models to determine the efficacy of an intervention is limited since individual animal models showed divergent results across interventions (Fig. 6.) and individual interventions showed divergent results across animal models (Fig. 7.). This is unsurprising for two main reasons: poor external validity of animal models and poor internal validity of animal studies.

4.1. Poor external validity of animal models for Alzheimer's disease

One cornerstone of assessing external validity of efficacy models is the extent to which animal models can simulate the disease. Human hallmarks of Alzheimer's disease include (age induced) cognitive impairment (Alzheimer's Association, 2018), decreased levels of neurotransmitters (Francis, 2005), increased neurological markers such as amyloid β plaques and tau tangles (Nisbet et al., 2015) or genetic alterations (Reitz, 2015).

A good animal model for Alzheimer's disease shows reproduction of histological, biochemical or systemic changes at cellular level, and changes at cognitive, functional or behavioral level (Götz et al., 2018). However, the multitude of possible targets for interventions, and the uncertainty around their level of involvement in the pathogenesis creates difficulties in selecting the right animal model. (Mullane and Williams, 2019; Reardon, 2018; Spinney, 2014).

Models identified in this study were chosen for modeling cognitive impairment (inclusion criterium) and one or more additional hallmarks of the human disease (Fig. 3 and S3 supplementary data). However, no justification was given on why one animal model was chosen over another.

We found 63 models, of which 91% was performed in mice and rats (Table 4 and Section 3.2.). The models were further categorized by the way in which cognitive impairment was acquired: 14% in spontaneous animal models, 57% in induced models, and 29% in genetically modified models. Since the models are a result of distinct biological processes, they all have limitations (Laurijssens et al., 2013).

The spontaneous models often lack the pathological features of the brain pathology like tau pathology, neurodegeneration or amyloid-

Table 4

Description of the animal models of efficacy for Alzheimer's disease showing cognitive impairment.

Model classification	Model ID	Model description
Spontaneous models	AGE	Aging induced
	DB/DB	Diabetic type 2 mouse, spontaneous mutation of *Lepr*^{db} gene
	F344	Accelerated aging, spontaneous
	KKAy	Diabetic type 2 mouse, spontaneous mutation of A^y gene
	SAMP8	Senescence Accelerated Mouse-Prone 8, spontaneous mutation
Induced models	192IgG-saporin	Immunotoxin 192 IgG-saporin induced
	2VO	Carotid occlusion induced
	AB	Amyloid β induced
	AF64A	AF64A induced; choline acetyltransferase inhibitor
	AGEprot	Advanced glycation end products induced
	AlCl3	Aluminum chloride induced
	AlCl3/Gal	Aluminum chloride and D-galactose induced
	AAV1-GFP	Adeno-associated virus vector-1-l1PP2A induced
	CH–Cu	Cholesterol and copper sulfate induced
	CMS	Chronic mild stress induced
	EE	Estradiol induced
	EtBD	Ethanol induced
	HFCD	High fat-cholesterol diet induced
	LESION	Physical lesion induced
	METH	Methamphetamine induced
	MK-801	Dizocilpine (MK-801) induced; NMDA receptor antagonist
	NMDA	N-methyl-d-aspartate induced
	OBX	Olfactory bulb extraction induced
	OKA	Okadaic acid induced
	OVX	Ovariectomy induced
	PTZ	Pentylenetetrazol induced
	QNB	3-quinyclidinyl benzilate induced
	SAPOVX	Immunotoxin 192 IgG-saporin and ovariectomy induced
	SCO	Scopolamine induced
	SD	Sleep deprivation induced
	STZ	Streptozotocin induced
	TBI	Traumatic brain injury induced
	TIF	Time induced topographical memory
	TOXO	*Toxoplasma gondii* infection induced
Genetically modified models	3xTg	*APP KM670/671NL (Swedish), MAPT P301L, PSEN1 M146V mutations*
	5XFAD	*APP KM670/671NL (Swedish), APP I716V (Florida), APP V717I (London), PSEN1 M146L (A > C), PSEN1 L286V mutations*
	AD11	*VKaD11HuCK, VHaD11HuCg carrying the light and heavy chain genes of the chimeric anti nerve growth factor antibody αD11*
	APOE-KO	*apoE knock out*
	APP/PS1	*APP KM670/671NL (Swedish), PSEN1: deltaE9 mutations*
	APP/PS1/COX	*APP KM670/67NL (Swedish), PSEN1: deltaE9 and COX-2 mutations*
	APP/PS1-21	*APP KM670/67NL (Swedish), PSEN1: L166P mutations*
	APP23	*APP KM670/671NL (Swedish) mutations*
	APPE693Q	*APP E693Q (Dutch) mutation*
	DGK	Diacylglycerol kinase β knockout
	J20	*APP KM670/671NL (Swedish), APP V717F (Indiana) mutations*
	Kir6.1HZ	Kir6.1 heterozygote mice; targeted *KCNJ8* disruption
	Kir6.2KO	Kir6.2 knockout mice; targeted *KCNJ8* disruption
	PDAPP	*APP V717F (Indiana) overexpression*
	Tg2576	*APP KM670/671NL (Swedish) mutation*
	TgCRND8	*APP KM670/671NL (Swedish), APP V717F (Indiana) mutation*
	Ts65Dn	aneuploid and carry the translocation chromosome, *Mmu17*^{16} *(17*^{16}*)*, as a freely segregating, supernumerary chromosome
	UAS-AB	*Aβ42 expression*

plaques. In induced models, the disease is often induced in healthy (young) animals, via local injections in the brain, resulting in a limited transmission of the pathogenesis throughout the body. Genetically modified models mainly express mutations found in relatively rare familial forms of Alzheimer's disease. (Bilkei-Gorzo, 2014; Dawson et al., 2018; Götz et al., 2018; Heuer et al., 2012).

Additionally, most animal models only allow the evaluation of a single hypothesis for Alzheimer disease. A few examples of mixed animal models have been reported, some of which used aged species in an induced model (Anand et al., 2017; Provensi et al., 2016; Zhai et al., 2016) or a genetically modified model, where cognitive impairment was tested in aged animals (Devi and Ohno, 2016; Martinez-Coria et al., 2010; Rodriguez-Rivera et al., 2011).

Another cornerstone of assessing the external validity is the response to treatment. We found that interventions with highest correlation (100% across models) or lowest correlation (0% across models) were mainly tested in a limited number of animal models (Fig. 7.).

Furthermore, the animal studies with "correct" outcomes were mainly a result of interventions of the cholinergic/glutaminergic class (the currently registered drugs) (Fig. 4B.). Additionally, the animal studies with "incorrect" outcome, used mainly animal models of the amyloid or aging hypothesis, which are not affected by these drugs (Fig. 5 and S4 supplementary data). Moreover, even in patients, only small beneficial effects were observed with these drugs, as shown by several systematic reviews (Birks et al., 2015; Birks and Harvey, 2018; Loy and Schneider, 2006; McShane et al., 2006). Unfortunately, there are currently no other drugs that offer overwhelming or even moderately positive effects in order to be included as a 'positive' control in animal models of efficacy for Alzheimer's disease.

The final cornerstone of the assessment of external validity is the outcome measure, which is cognition in Alzheimer's disease. Human cognition includes processes of perception, attention, learning, short-term and long-term memory, speech and social cognition, of which the profile can be determined with test batteries such as the Cambridge
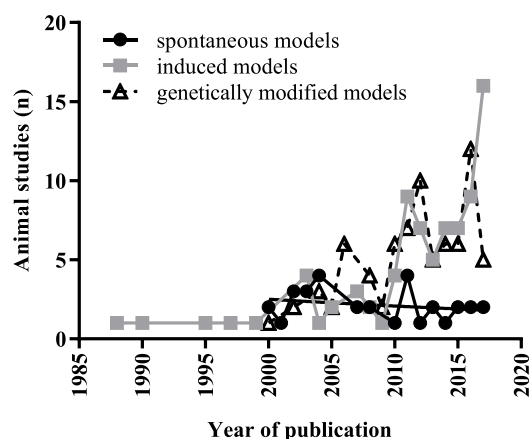
Fig. 2. Evolution of the amount of animal studies using different animal model classes between 1988 and 2017.

Total = 208 animal studies

- ■ model of ageing-related cognitive impairment
- ■ model showing hallmarks of Alzheimer's disease
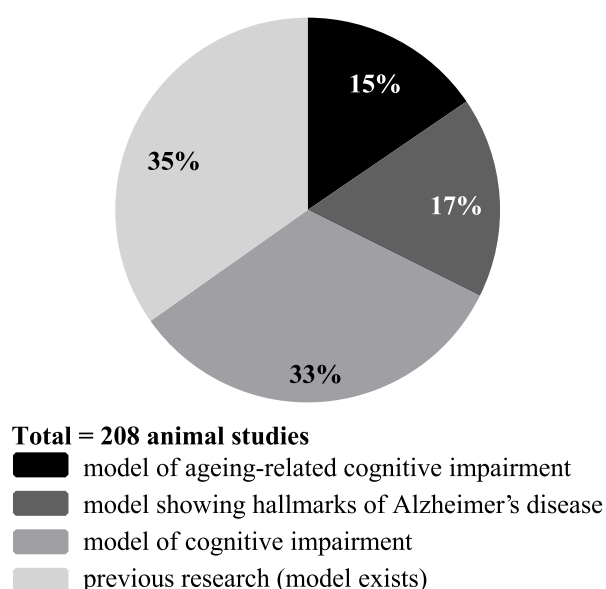- ■ model of cognitive impairment
- ☐ previous research (model exists)

Fig. 3. Author's reported motivation for their choice of animal model of efficacy for Alzheimer's disease.

Neuropsychological Test Automated Battery (CANTAB). However, the individual cognitive impairment profile is highly dependent on the disease stage as well as existing comorbidities. (Millan et al., 2012; Robbins et al., 1994). This differentiation in disease severity is not reflected in animal models, since mostly only single outcome measures are used. Additionally, the validation of cognitive tasks is still a topic of debate (Decker, 2006; van der Staay et al., 2009). Therefore, the choice

Fig. 5. Percentage of animal studies with of cholinergic/glutaminergic interventions (the current registered drugs), with correct (blue), unclear (orange) or incorrect (red) predictions for their clinical outcome in animal models for Alzheimer's disease. Data are grouped per animal model class: all animal models (ALL), spontaneous models (SP), induced models (IN) or genetically modified models (GM).

of an animal species (with homology in brain physiology and function) as well as an appropriate cognitive test (testing disease relevant cognitive impairment) in animals is challenging (Keeler and Robbins, 2011). In our study, we found 16 different methods which all assessed different aspects of learning and memory in animals (Fig. 8.). How and to what extent animal and human cognition relate is unclear, especially because unlike human cognitive tests, many animal methods include a positive or negative reinforcement. To our knowledge no systematic validation and back-translation of human and animal cognitive tests exists (Al Dahhan et al., 2019), except in the field of Schizrenia (Lustig et al., 2013; Young and Light, 2018) and Psychiatry (Halcomb et al., 2019).

4.2. Poor internal validity of animal studies for Alzheimer's disease

Besides the variation in available methods to assess cognition in animals, the study outcome may also be affected by variations in how the test is carried out. This is exemplified by variations in pool size, water temperature and number of training days in a version of the Morris Water Maze, which was used most frequently in our study. (Egan et al., 2016; Stukalin and Einat, 2019).

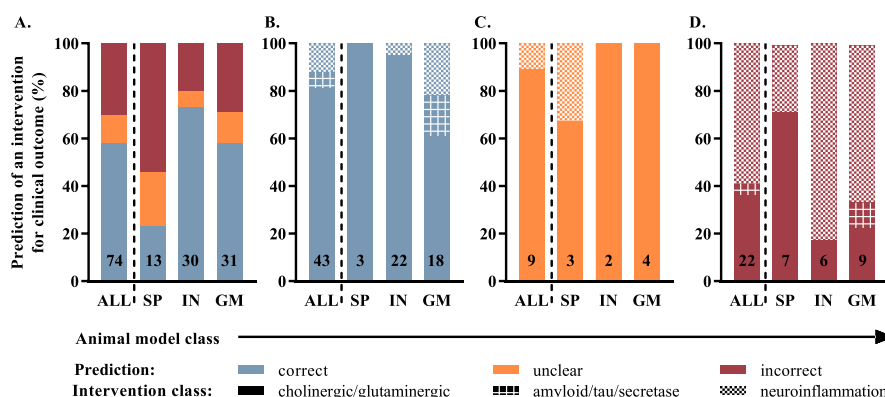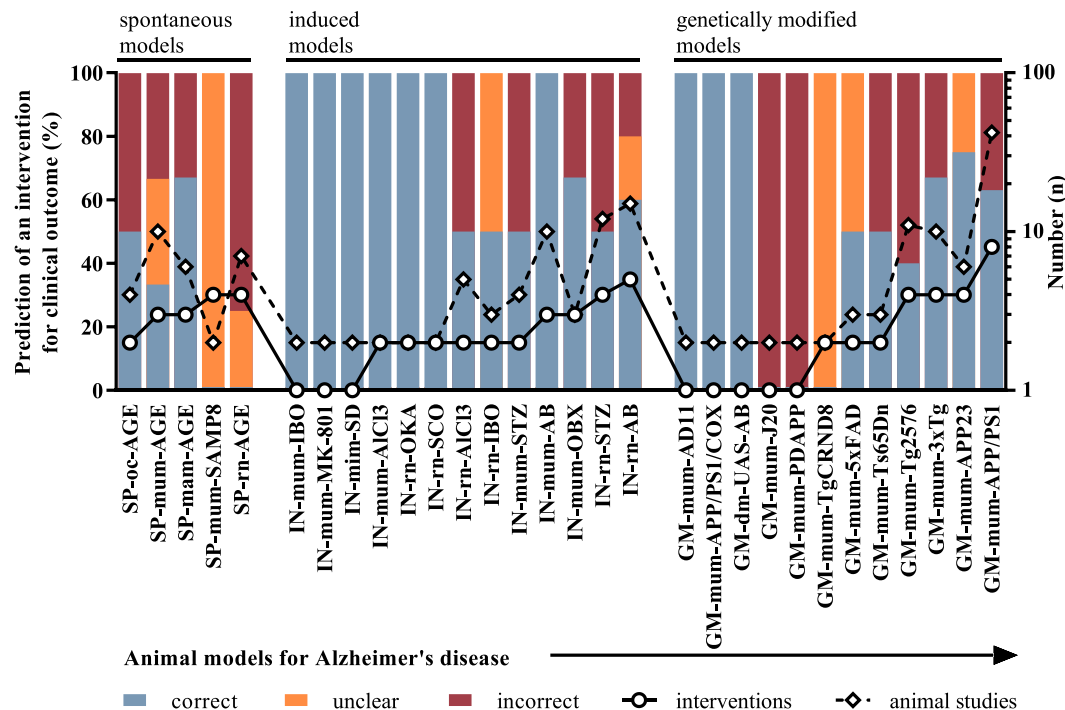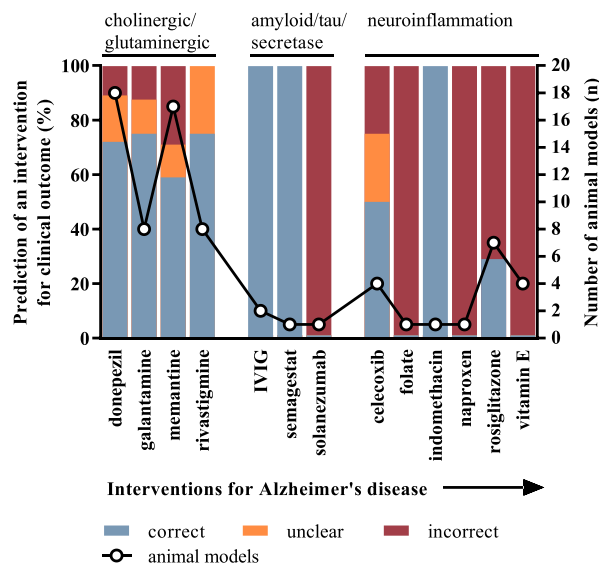Furthermore, we also found poor (maximal 65%) compliance with

Fig. 4. Percentage of interventions with correct (blue), unclear (orange) or incorrect (red) predictions for their clinical outcome in animal models for Alzheimer's disease (panel A) as well as the percentage of cholinergic/glutaminergic (closed bar pattern), amyloid/tau/secretase (squared bar pattern), neuroinflammation (diamond bar pattern) intervention classes which predicted their clinical outcome correct (panel B), unclear (panel C) or incorrect (panel D) in these models. Data are grouped per animal model class: all animal models (ALL), spontaneous models (SP), induced models (IN) or genetically modified models (GM). Values in the bars show the number of intervention-model combinations.
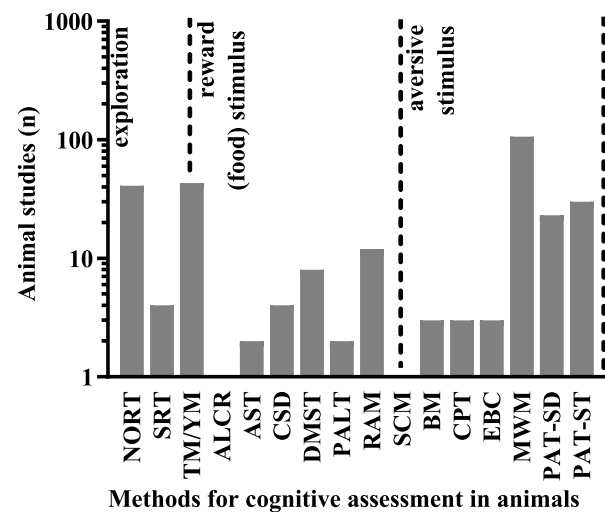
Fig. 6. Percentage (left Y-axis) and number (right Y-axis) of animal studies (dotted line) or interventions (straight line) with correct (blue bars), unclear (orange bars) or incorrect (red bars) predictions for their clinical outcome in an animal models for Alzheimer's disease. Data are grouped per animal model class: all animal models (ALL), spontaneous models (SP), induced models (IN) or genetically modified models (GM). Animal models were carried out in different species; *Drosophila megalogaster (dm)*, *Macaca mulatta (mam)*, *Mus musculus (mum)*, *Orycolagus cuniculus (oc)* or *Rattus norvegicus (rn)*. Animal model name abbreviations are explained in Table 4.

Fig. 7. Percentage (left Y-axis) and number of animal models for Alzheimer's disease (right Y-axis) with correct (blue bars), unclear (orange bars) or incorrect (red bars) predictions for their clinical outcome. Data are grouped per intervention class: cholinergic/glutaminergic, amyloid/tau/secretase or neuroinflammation.

the ARRIVE guidelines (Fig. 9.). Despite the introduction of reporting guidelines and an increasing number of publications creating awareness of how poor design affects study outcomes (Henderson et al., 2015; Hirst et al., 2014; Hooijmans and Ritskes-Hoitinga, 2013), the reporting quality of animal studies did not improve between 1988 and 2018. As a result, animal models, may seem better than they are in practice.
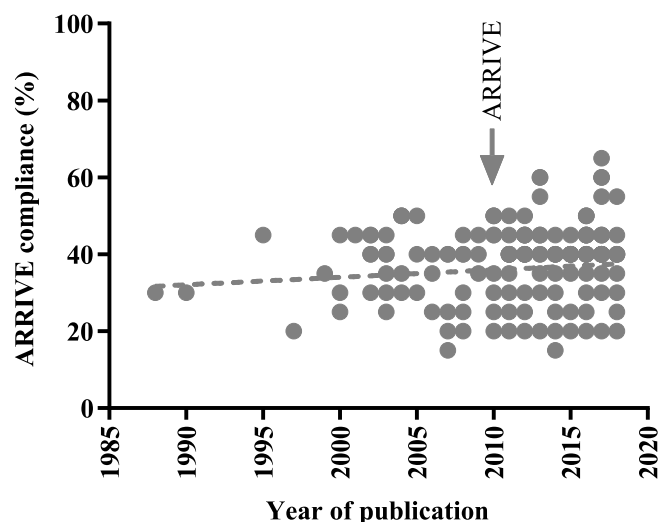

Fig. 8. Differently motivated outcome measures for cognition. Data is shown as number of studies using a specific outcome measure. Natural exploratory behavior: Novel Object Recognition Task (NORT), Spatial Recognition Test (SRT), T- or Y-maze Alteration Test (TM/YM). Reward (food) motivated: T- or Y-maze Alteration Test (TM/YM), Alternating Lever Cycle-Ratio Schedule test (ALCR), Attention Set-Shifting Test (AST), Contextual Serial Discrimination Task (CSD), Delayed Matching-to-Sample/Position Task (DMTS), Paired Associate Learning Task (PALT), Radial Arm Maze (RAM). Aversive stimulus motivated: Contextual Serial Discrimination Task (CSD), Barnes Maze (BM), Circular Platform Task (CPT), Eyeblink Conditioning Procedure (EBC), Morris Water Maze (MWM), Passive Avoidance Step-Down Test (PAT-SD), Passive Avoidance Step-Through Test (PAT-ST) and Stone Complex Maze (SCM). Detailed information on the methods is available in Supplementary data S2.

Table 5
Author's disclosure on different aspects of the animal study population, housing and husbandry as well as statistics, blinding and randomization. Included were author's reporting on these items as such. Item reporting was not evaluated on level of correctness.

Animal study aspects	Items	Item disclosed (% of animal studies)
Population	strain	99
	sex	88
	age and/or weight	99
Housing and husbandry	facility type	6
	cage type	7
	bedding type	6
	number of cage companions	39
	light/dark cycle	68
	temperature and/or humidity	54
	quality of water	8
	food type	18
	access to food and water	72
	environmental enrichment	2
	acclimatization period	29
Statistics, blinding and randomization	sample size	92
	sample size calculation	1
	allocation concealment	3
	blinded outcome assessment	9
	random allocation to cage	1
	random allocation to treatment	24

Fig. 9. Reporting quality assessment of animal studies for Alzheimer's disease between 1998 and 2018. The ARRIVE guideline was introduced in 2010.

4.3. Limitations of our study

Besides the poor reporting quality, another limitation of our study is the inclusion of clinical trials with published results including statistics either in the clinicaltrials.gov database or on PubMed. The targeted search in PubMed used the NCT number [All Fields]. Some trials might have been missed, if this number was not reported in the fields searched by PubMed.

Next, we did not consider the effect of dosing, formulation and duration of the intervention on the outcomes. Additionally, we reported the overall outcome ("improved" or "no effect") of an intervention in a specific animal model and not the actual effect size. Actual effect sizes, including exposure data, when combined for each intervention, could

provide a better estimation of the true effect. Furthermore, we assumed the animal studies were of equal value, which is not necessarily the case due to differences in design and methodology. In case of an equal amount of "improved" and "no effect" animal studies for the same intervention, the overall outcome was listed as "mixed" despite the fact that we know that part of the negative results found in animal models are not published (Fanelli, 2012) and these "mixed" outcomes are more likely to be a "no effect" outcome.

Clinical trials are the gold standard, while there is mounting evidence that patients with Alzheimer's disease are very heterogeneous and the patient population included in previous trials, needs further stratification. This may affect the design of future clinical trials (Au et al., 2015; Golde et al., 2018).

4.4. Implications for drug development

The failure of animal models of efficacy for Alzheimer's disease to correlate with the effect in patients is known in the field and several reviews have discussed this issue (Al Dahhan et al., 2019; Goodarzi et al., 2019; Mullane and Williams, 2019; Quinn, 2018). However, for the individual animal studies cited in these reviews, it is often unclear which interventions and outcomes were compared between animal and human. To our knowledge, the only paper with a broad systematic approach was the study of Zeiss (2015). New in our study is that we specifically chose a clinical efficacy outcome marker and identified all relevant animal studies via a targeted search. We assumed complete face validity of the animal models, i.e. are the same symptoms present in the model and the human disease, since we only included studies where cognitive impairment was observed and measured. We assumed limited construct validity: as mainly single hypotheses were tested in models of a multifactorial disease. And we evaluated the predictive validity, i.e. whether the same interventions are effective or ineffective in the model when compared to the clinic. In that sense, we developed an approach that can be used across different indications for that purpose, using the Alzheimer's disease animal models of efficacy as a case study.

Our study shows that the 'convenient' choice of an animal model for Alzheimer disease, e.g. the model that has been used before or the model that is most often used, might not be the optimal approach. When selecting (or developing) an animal model, it is impossible to simulate all disease phenomena in one model. Moreover, the ability to interpret the effect of an intervention on the (induced) phenomena a model should be kept as simple as possible (Keeler and Robbins, 2011).

Suggestions to improve (or develop new) animal models of efficacy for Alzheimer's disease:

(1) Create a translational profile of the model with use of a recently developed tool such as the Framework to Identify Models of Disease (FIMD) (Ferreira et al., 2019). This approach can also be used for new models since it indicates which aspects must be investigated to validate an animal model.
(2) Validate the model systematically, e.g. evaluation of the effects of different rearing and housing conditions, sex, age, outcome measures and eventually validate the model across species (van der Staay et al., 2009). Include (at least the) outcome markers of clinical efficacy markers.
(3) Design, register, execute and report the animal study properly (using standards such as PREPARE, ARRIVE, preclinicaltrials.eu). Compliance must be verified by funding agencies and editors.
(4) Share all raw data obtained in animal studies and clinical trials; this will facilitate re-use/analysis and promotes learning. Create platforms where these data can be combined as suggested by (Alteri and Guizzaro, 2018).

Similarly, these suggestions also apply to all *ex-vivo* and *in vitro* experiments to determine the efficacy of interventions for Alzheimer's

disease.

For complex diseases such as Alzheimer's disease one could consider a (properly designed) test battery of different (animal) models simulating different aspects of the disease (Stukalin and Einat, 2019; Webster et al., 2014; Wolf et al., 2016). Additionally, the often-used words "proof-of-concept" or "proof-of-principle" in a multifactorial disease, such as Alzheimer's disease, are in reality the proof of a specific hypothesis tested. Therefore, besides general categorization of an animal model by the way the disease is acquired (spontaneous, induced or genetically modified), further categorization of animal models based on hypothesis testing may facilitate easier pattern recognition and back-translation to what is found in the Alzheimer patient (Mullane and Williams, 2019).

5. Conclusion

Currently, no animal models can be recommended for determining the efficacy of interventions in Alzheimer's disease. Investigators, ethical committees and funding bodies should be reminded that, in the way animal models are currently used to test the efficacy of novel drugs in the field of Alzheimer's disease, these models are not useful to predict clinical efficacy. There is an urgent need for new interventions, validated (and/or other) models (both *in vivo* and *in vitro*), in which translational outcomes are measured and animal studies are properly designed and reported.

Conflicts of interest

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: GSF reports personal fees from Merck KGaA and SDD Consulting B.V. outside of the submitted work. DVG reports personal fees from Nutricia Research B.V. outside of the submitted work. None of the other authors has any conflicts of interest.

Authors' contributions

Conceptualization by DHVG and PJKM; Data curation by DHVG and GSF; Formal analysis by DHVG; Funding acquisition by HS; Investigation by DHVG; Methodology by DHVG and GSF; Supervision by PJKM; Visualization by DHVG; Writing - original draft by DHVG; Writing - review & editing by DHVG, GSF, PJKM, WB, CCGW, EM and HS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejphar.2019.172524.

References

Al Dahhan, N.Z., De Felice, F.G., Munoz, D.P., 2019. Potentials and pitfalls of cross-translational models of cognitive impairment. Front. Behav. Neurosci. 13, 48.

Alteri, E., Guizzaro, L., 2018. Be open about drug failures to speed up research. Nature 563, 317–319.

Alzheimer's Association, 2018. 2018 Alzheimer's disease facts and figures. Alzheimer's Dementia 14, 367–429.

Anand, A., Patience, A.A., Sharma, N., Khurana, N., 2017. The present and future of pharmacotherapy of Alzheimer's disease: a comprehensive review. Eur. J. Pharmacol. 815, 364–375.

Au, R., Piers, R.J., Lancashire, L., 2015. Back to the future: Alzheimer's disease heterogeneity revisited. Alzheimers Dement (Amst) 1, 368–370.

Begley, C.G., Ioannidis, J.P., 2015. Reproducibility in science: improving the standard for basic and preclinical research. Circ. Res. 116, 116–126.

Bilkei-Gorzo, A., 2014. Genetic mouse models of brain ageing and Alzheimer's disease. Pharmacol. Ther. 142, 244–257.

Birks, J.S., Harvey, R.J., 2018. Donepezil for dementia due to Alzheimer's disease. Cochrane Database Syst. Rev. 6.

Birks, J.S., Chong, L.Y., Grimley Evans, J., 2015. Rivastigmine for Alzheimer's disease. Cochrane Database Syst. Rev. 9.

Cummings, J.L., Morstorf, T., Zhong, K., 2014. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. Alzheimer's Res. Ther. 6, 37.

Dawson, T.M., Golde, T.E., Lagier-Tourenne, C., 2018. Animal models of neurodegenerative diseases. Nat. Neurosci. 21, 1370–1379.

Decker, M.W., 2006. Cognition models and drug discovery. In: Levin, E.D., Buccafusco, J.J. (Eds.), Animal Models of Cognitive Impairment, Boca Raton (FL).

Dekker, M., Bouvy, J.C., O'Rourke, D., Thompson, R., Makady, A., Jonsson, P., Gispen-de Wied, C.C., 2019. Alignment of european regulatory and Health technology assessments: a review of licensed products for Alzheimer's disease. Front. Med. 6, 73.

Devi, L., Ohno, M., 2016. Cognitive benefits of memantine in Alzheimer's 5XFAD model mice decline during advanced disease stages. Pharmacol. Biochem. Behav. 144, 60–66.

Egan, K.J., Vesterinen, H.M., Beglopoulos, V., Sena, E.S., Macleod, M.R., 2016. From a mouse: systematic analysis reveals limitations of experiments testing interventions in Alzheimer's disease mouse models. Evid. Based Prev. Med. 3, e00015.

European Medicines Agency, 2004. European Public Assessment Report on Memantine. . https://www.ema.europa.eu/documents/scientific-discussion/ebixa-epar-scientific-discussion_en.pdf, Accessed date: 28 May 2018.

European Medicines Agency, 2005. European Public Assessment Report on Rivastigmine. . https://www.ema.europa.eu/documents/scientific-discussion/exelon-epar-scientific-discussion_en.pdf, Accessed date: 28 May 2018.

European Medicines Agency, 2018. Guideline on the Clinical Investigation of Medicines for the Treatment of Alzheimer's Disease. https://www.ema.europa.eu/documents/scientific-guideline/guideline-clinical-investigation-medicines-treatment-alzheimers-disease-revision-2_en.pdf, Accessed date: 28 May 2018.

Fanelli, D., 2012. Negative results are disappearing from most disciplines and countries. Scientometrics 90, 891–904.

Ferreira, G.S., Veening-Griffioen, D.H., Boon, W.P.C., Moors, E.H.M., Gispen-de Wied, C.C., Schellekens, H., van Meer, P.J.K., 2019. A standardised framework to identify optimal animal models for efficacy assessment in drug development. PLoS One 14, e0218014.

Francis, P.T., 2005. The interplay of neurotransmitters in Alzheimer's disease. CNS Spectr. 10, 6–9.

Golde, T.E., DeKosky, S.T., Galasko, D., 2018. Alzheimer's disease: the right drug, the right time. Science 362, 1250–1251.

Goodarzi, P., Payab, M., Alavi-Moghadam, S., Larijani, B., Rahim, F., Bana, N., Sarvari, M., Adibi, H., Foroughi Heravani, N., Hadavandkhani, M., Arjmand, B., 2019. Development and validation of Alzheimer's disease animal model for the purpose of regenerative medicine. Cell Tissue Bank. 20, 141–151.

Götz, J., Bodea, L.-G., Goedert, M., 2018. Rodent models for Alzheimer disease. Nat. Rev. Neurosci. 19, 583–598.

Halcomb, M., Argyriou, E., Cyders, M.A., 2019. Integrating preclinical and clinical models of negative urgency. Front. Psychiatry 10, 324.

Hartung, D.M., Zarin, D.A., Guise, J.M., McDonagh, M., Paynter, R., Helfand, M., 2014. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. Ann. Intern. Med. 160, 477–483.

Hau, J., 2008. Animal models for human diseases. In: Conn, P.M. (Ed.), Sourcebook of Models for Biomedical Research. Humana Press, Totowa, NJ, pp. 3–8.

Henderson, V.C., Demko, N., Hakala, A., MacKinnon, N., Federico, C.A., Fergusson, D., Kimmelman, J., 2015. A meta-analysis of threats to valid clinical inference in preclinical research of sunitinib. Elife 4, e08351.

Heuer, E., Rosen, R.F., Cintron, A., Walker, L.C., 2012. Nonhuman primate models of Alzheimer-like cerebral proteopathy. Curr. Pharmaceut. Des. 18, 1159–1169.

Hirst, J.A., Howick, J., Aronson, J.K., Roberts, N., Perera, R., Koshiaris, C., Heneghan, C., 2014. The need for randomization in animal trials: an overview of systematic reviews. PLoS One 9, e98856.

Hooijmans, C.R., Ritskes-Hoitinga, M., 2013. Progress in using systematic reviews of animal studies to improve translational research. PLoS Med. 10, e1001482.

Hooijmans, C.R., Tillema, A., Leenaars, M., Ritskes-Hoitinga, M., 2010. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. Lab. Anim. 44, 170–175.

Howells, D.W., Sena, E.S., Macleod, M.R., 2014. Bringing rigour to translational medicine. Nat. Rev. Neurol. 10, 37–43.

Ioannidis, J.P., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R., Moher, D., Schulz, K.F., Tibshirani, R., 2014. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 383, 166–175.

Keeler, J.F., Robbins, T.W., 2011. Translating cognition from animals to humans. Biochem. Pharmacol. 81, 1356–1366.

Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M., Altman, D.G., 2010. Improving bioscience research reporting: the ARRIVE guidelines for animal research. PLoS Biol. 8, e1000412.

Laurijssens, B., Aujard, F., Rahman, A., 2013. Animal models of Alzheimer's disease and drug development. Drug Discov. Today Technol. 10, e319–327.

Leenaars, M., Hooijmans, C.R., van Veggel, N., ter Riet, G., Leeflang, M., Hooft, L., van der Wilt, G.J., Tillema, A., Ritskes-Hoitinga, M., 2012. A step-by-step guide to systematically identify all relevant animal studies. Lab. Anim. 46, 24–31.

Loy, C., Schneider, L., 2006. Galantamine for Alzheimer's disease and mild cognitive impairment. Cochrane Database Syst. Rev. 1.

Lustig, C., Kozak, R., Sarter, M., Young, J.W., Robbins, T.W., 2013. CNTRICS final animal model task selection: control of attention. Neurosci. Biobehav. Rev. 37, 2099–2110.

Martinez-Coria, H., Green, K.N., Billings, L.M., Kitazawa, M., Albrecht, M., Rammes, G., Parsons, C.G., Gupta, S., Banerjee, P., LaFerla, F.M., 2010. Memantine improves cognition and reduces Alzheimer's-like neuropathology in transgenic mice. Am. J. Pathol. 176, 870–880.

McKinney Jr., W.T., Bunney Jr., W.E., 1969. Animal model of depression. I. Review of evidence: implications for research. Arch. Gen. Psychiatr. 21, 240–248.

McShane, R., Areosa Sastre, A., Minakaran, N., 2006. Memantine for dementia. Cochrane Database Syst. Rev. 2.

Millan, M.J., Agid, Y., Brune, M., Bullmore, E.T., Carter, C.S., Clayton, N.S., Connor, R., Davis, S., Deakin, B., DeRubeis, R.J., Dubois, B., Geyer, M.A., Goodwin, G.M., Gorwood, P., Jay, T.M., Joels, M., Mansuy, I.M., Meyer-Lindenberg, A., Murphy, D., Rolls, E., Saletu, B., Spedding, M., Sweeney, J., Whittington, M., Young, L.J., 2012. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. Nat. Rev. Drug Discov. 11, 141–168.

Mullane, K., Williams, M., 2019. Preclinical models of Alzheimer's disease: relevance and translational validity. Curr Protoc Pharmacol 84, e57.

Nisbet, R.M., Polanco, J.C., Ittner, L.M., Gotz, J., 2015. Tau aggregation and its interplay with amyloid-beta. Acta Neuropathol. 129, 207–220.

Provensi, G., Costa, A., Passani, M.B., Blandina, P., 2016. Donepezil, an acetylcholine esterase inhibitor, and ABT-239, a histamine H3 receptor antagonist/inverse agonist, require the integrity of brain histamine system to exert biochemical and procognitive effects in the mouse. Neuropharmacology 109, 139–147.

Quinn, J.F., 2018. Lost in translation? Finding our way to effective Alzheimer's disease therapies. J. Alzheimer's Dis. : JAD 64, S33–S39.

Reardon, S., 2018. Alzheimer's researchers seek better mice. Nature 563, 611–612.

Reitz, C., 2015. Genetic diagnosis and prognosis of Alzheimer's disease: challenges and opportunities. Expert Rev. Mol. Diagn 15, 339–348.

Rinaldi, A., 2018. Setbacks and promises for drugs against Alzheimer's disease: as pharmaceutical companies are retreating from drug development for Alzheimer's, new approaches are being tested in academia and biotech companies. EMBO Rep. 19.

Ritskes-Hoitinga, M., Wever, K., 2018. Improving the conduct, reporting, and appraisal of animal research. BMJ 360, j4935.

Robbins, T.W., James, M., Owen, A.M., Sahakian, B.J., McInnes, L., Rabbitt, P., 1994. Cambridge neuropsychological test automated battery (CANTAB): a factor Analytic study of a large sample of normal elderly volunteers. Dement. Geriatr. Cognit. Disord. 5, 266–281.

Rodriguez-Rivera, J., Denner, L., Dineley, K.T., 2011. Rosiglitazone reversal of Tg2576 cognitive deficits is independent of peripheral gluco-regulatory status. Behav. Brain Res. 216, 255–261.

Sams-Dodd, F., 2006. Strategies to optimize the validity of disease models in the drug discovery process. Drug Discov. Today 11, 355–363.

Smith, A.J., Clutton, R.E., Lilley, E., Hansen, K.E.A., Brattelid, T., 2018. PREPARE: guidelines for planning animal research and testing. Lab. Anim. 52, 135–141.

Spinney, L., 2014. Alzheimer's disease: the forgetting gene. Nature 510, 26–28.

Stukalin, Y., Einat, H., 2019. Analyzing test batteries in animal models of psychopathology with multivariate analysis of variance (MANOVA): one possible approach to increase external validity. Pharmacol. Biochem. Behav. 178, 51–55.

United States Food and Drug Administration, 2001. FDA Approved Drug Products Review on Galantamine. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2001/21-169_Reminyl_biopharmr.pdf, Accessed date: 28 May 2018.

United States Food and Drug Administration, 2006. FDA Approved Drug Products Review on Donepezil. https://www.accessdata.fda.gov/drugsatfda_docs/label/2006/020690s026,021720s003lbl.pdf, Accessed date: 28 May 2018.

van der Staay, F.J., Arndt, S.S., Nordquist, R.E., 2009. Evaluation of animal models of neurobehavioral disorders. Behav. Brain Funct. 5, 11.

van der Worp, H.B., Howells, D.W., Sena, E.S., Porritt, M.J., Rewell, S., O'Collins, V., Macleod, M.R., 2010. Can animal models of disease reliably inform human studies? PLoS Med. 7, e1000245.

van Meer, P.J.K., Graham, M.L., Schuurman, H.-J., 2015. The safety, efficacy and regulatory triangle in drug development: impact for animal models and the use of animals. Eur. J. Pharmacol. 759, 3–13.

Webster, S.J., Bachstetter, A.D., Nelson, P.T., Schmitt, F.A., Van Eldik, L.J., 2014. Using mice to model Alzheimer's dementia: an overview of the clinical disease and the preclinical behavioral changes in 10 mouse models. Front. Genet. 5, 88.

Willner, P., 1984. The validity of animal models of depression. Psychopharmacology (Berlin) 83, 1–16.

Wolf, A., Bauer, B., Abner, E.L., Ashkenazy-Frolinger, T., Hartz, A.M., 2016. A comprehensive behavioral test battery to assess learning and memory in 129S6/tg2576 mice. PLoS One 11, e0147733.

Young, J.W., Light, G.A., 2018. Cross-species neurophysiological biomarkers of attentional dysfunction in Schizophrenia: bridging the translational gap. Neuropsychopharmacology 43, 230–231.

Zeiss, C.J., 2015. Improving the predictive value of interventional animal models data. Drug Discov. Today 20, 475–482.

Zhai, Y., Yamashita, T., Nakano, Y., Sun, Z., Shang, J., Feng, T., Morihara, R., Fukui, Y., Ohta, Y., Hishikawa, N., Abe, K., 2016. Chronic cerebral hypoperfusion accelerates Alzheimer's disease pathology with cerebrovascular remodeling in a novel mouse model. J. Alzheimer's Dis. 53, 893–905.