Tools to enable animal to human translation:

assessing the value of disease models

Guilherme Sant'Anna Ferreira

Tools to enable animal to human translation: assessing the value of disease models

Guilherme Sant'Anna Ferreira

Tools to enable animal to human translation:
assessing the value of disease models

Instrumenten die de vertaling van dier naar mens mogelijk
maken: beoordeling van de waarde van ziekte modellen
(met een samenvatting in het Nederlands)

Ferramentas para possibilitar a tradução de animais a humanos:
avaliando o valor de modelos de doença
(com resumo em Português)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 22 januari 2021 des middags te 2.30 uur

door

Guilherme Sant'Anna Ferreira

geboren op 22 april 1991
te Rio de Janeiro, Brazilië

Promotoren:
Prof. dr. H. Schellekens
Prof. dr. E.H.M. Moors

Copromotoren:
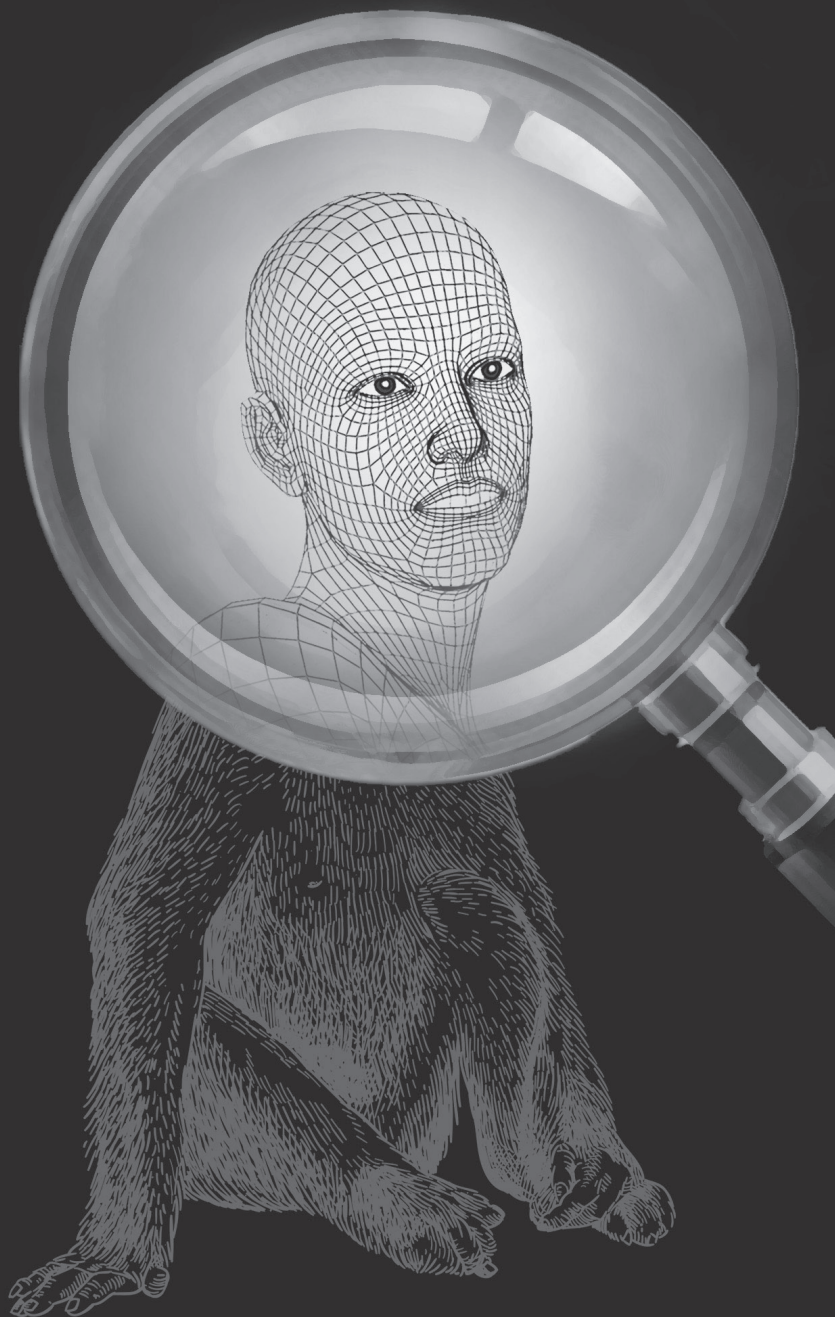Dr. P.J.K. Van Meer
Dr. W.P.C. Boon

"All models are wrong, but some models are useful."
George Box

Table of contents

1

GENERAL INTRODUCTION

The journey from bench to bedside is paved with failures as most drugs getting into clinical development will never reach the patients they are meant to treat (1–3). The rate of failure, also known as attrition rate, has been a point of discussion for more than a decade (1). Differently from the 1990s, when 40% of drugs failed because of pharmacokinetics and bioavailability issues, now the primary reasons for failure are clinical efficacy, safety and commercial, in this order (1–3). Since new drugs frequently advance into the clinic based on animal studies, insufficient efficacy or safety in humans suggests that animal models are not predictive enough (1,4,5). On these grounds, animal testing is now under increased scrutiny as drug developers search for potential explanations and solutions (4,6,7).

A closer look at animal research sheds light on an important factor for this lack of predictivity: animal experiments are seldom reproducible. Two landmark papers on the reproducibility of animal research published worrying results. Prinz and colleagues tried to reproduce in-house, at Bayer Healthcare, the results of 67 publications in oncology, women's health and cardiovascular diseases (8). They only successfully reproduced 21% of the studies, with an additional 7% being partially reproducible. The following year, Begley and Ellis published a similar analysis, conducted at Amgen, with 53 'landmark papers' (9). The results confirmed Prinz's: only 6% of studies were reproduced.

With such low rates of reproducibility, some clinical trials may be advancing based on spurious results. Nevertheless, what causes results from an animal experiment not to translate to the clinic? Two concepts are essential to answer this question: internal and external validities. The internal validity refers to whether the findings of an experiment can be attributed to the differences between groups and not to random chance. In contrast, external validity refers to the generalisation of results to different settings, such as animal to human translation (4).

The effects of poor internal validity have already been explored in clinical and preclinical studies (10,11). In animals, studies that did not mention randomisation or blinding – two measures that address most types of bias – were more than 3 times more likely to report positive outcomes. Vogt and colleagues showed that measures to prevent bias were barely described in animal experiments applications and their corresponding publications – with none reporting the sample size calculation (12). Moreover, these findings seem to be commonplace also in high-esteemed institutions and high-impact journals (13).

The ubiquity of suboptimal study design resulted in many initiatives to address internal validity. The Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines, published in 2010, were the first harmonised reporting guidelines across research areas. With harmonised reporting standards, it is clearer how reliable the

results from a given experiment are. Nonetheless, despite the endorsement by over 300 journals, translation to eight languages and extensive divulgation, the ARRIVE guidelines have not yet improved the reporting quality in animal research (14,15).

Next to the ARRIVE guidelines, the Planning Research and Experimental Procedures on Animals: Recommendations for Excellence (PREPARE) guidelines, published in 2018, address the study design and planning aspects of animal studies (16). The European Quality in Preclinical Data (EQPID), a project from the Innovative Medicines Initiative (IMI) launched in 2017, focuses on strengthening the robustness, rigour and validity of preclinical data (17). EQIPD's publications explore a range of topics and include a handbook of good preclinical research practices. Thus, an improvement of the internal validity now depends on the implementation, rather than development, of guidelines and recommendations.

For external validity, some aspects of study design are also fundamental. For instance, one of the reasons often mentioned for the lack of translatability in acute stroke research is the time to intervention (18). In animals, most neuroprotective agents are administered before or shortly after stroke induction while in humans, there is usually a delay of at least 6 hours. However, a sound study design alone does not address a significant aspect of external validity: the animal models themselves. A telling example of the importance of animal model selection can be found in research in Amyotrophic Lateral Sclerosis (ALS). Whereas most studies use the superoxide dismutase 1 (SOD1) mouse model, the homologous human gene is present in only 10% of patients of what is called 'familial' ALS (19). Also, there are significant sex differences in time to onset that can affect survival endpoints directly. The abysmal success rates of ALS drugs in the clinic can be at least partially attributed to treating 'sporadic' ALS patients with drugs tested in a 'familial' ALS mouse model (19,20). Hence, how can we evaluate the external validity of animal models of human conditions, also known as animal models of disease or disease models?

Traditionally, three concepts have been used to assess disease models regarding their ability to simulate the human condition: face, construct and predictive validities. These concepts were first introduced for animal models of depression by McKinney and Bunney and later popularised by Willner (21,22). The face validity refers to how the disease is manifested in the animal when compared to humans, e.g. symptomatology; construct validity refers to whether the human and animal conditions share the underlying pathophysiology; predictive validity refers to whether animals can replicate the human response to effective drugs. Although these concepts present the first set of criteria to assess external validity, they are not standardised, may be interpreted differently by researchers and preclude an easy comparison between models.

General Introduction

A tool developed by Sams-Dodd in 2006 and further elaborated by Denayer and colleagues in 2014, resolved some of the limitations related to the concepts of validity (23,24). The tool applies to *in vitro* and *in vivo* models and is based on the 'distance' (including a scoring system) to the patient regarding the species, disease simulation, face validity, complexity and predictivity. This tool represents an advancement in terms of standardisation, but it lacks a more in-depth analysis of pathophysiological features, and it may lead to two models scoring the same with few possibilities for differentiation.

Currently, the validity concepts and Sams-Dodd's tool do not offer a biologically encompassing, systematic and evidence-based method to evaluate disease models' external validity. Nonetheless, it is only by establishing the essential information through which animal models can be evaluated that we can select models more likely to predict the human response. Hence, there is a need for new ways to discriminate among potentially predictive animal models of disease.

Aim and Outline of the Thesis

This thesis aims to fill the gap of producing new ways to discriminate among potentially predictive animal models. Here, we define the standards to guide the assessment, comparison and validation of animal models used to test the preliminary efficacy of new drugs. We base our approach on outlining the relationships between animal and human disease at physiological, pharmacological and therapeutic levels. To this end, we present a framework to evaluate the similarities between disease models and patients, delineating a clear procedure to validate disease models on these grounds. Furthermore, we complement our framework by applying existing methodologies to investigate the predictivity of animal models regarding pharmacological and therapeutic effects. When combined, these methods provide a holistic, systematic and evidence-based manner to select disease models based on their ability to predict the human response.

In **Chapter 2**, we present the development of the Framework to Identify Models of Disease (FIMD). We detail the rationale behind each feature of FIMD, and provide instructions for its application in the assessment, validation and comparison of animal models of disease. In **Chapter 3**, we discuss the validation of FIMD, which included both a simplified version of the framework (pilot study) and a full validation of two animal models of Type 2 Diabetes (T2D) and Duchenne Muscular Dystrophy (DMD), respectively. We elaborate on the application of all features of FIMD, providing a stepwise approach for its implementation.

FIMD assesses most aspects relevant to the evaluation of disease models' external validity at (patho)physiological level, but only superficially explores the concordance between animal and human drug response. In **Chapter 4**, we further explore the results from FIMD's pilot study in T2D, which showed only minor differences between the two assessed models. Therefore, we conduct a systematic review and meta-analysis of animal studies to validate the pilot study results at the therapeutic level (glucose-lowering effect). Moreover, we establish the basis for the application of systematic reviews and meta-analysis to compare animal and human efficacy data.

In **Chapter 5**, we explore preclinical pharmacokinetic and pharmacodynamic predictions in early clinical development to assess disease models at the pharmacological level. To this goal, Investigator's Brochures (IBs), a critical document that summarises all relevant information before first-in-human studies, are used as a source material. We use the IB-derisk tool (developed at the Centre for Human Drug Research – CHDR, Leiden, NL) to integrate animal and human pharmacokinetic and pharmacodynamic data. The preclinical data in the IB is compared with early clinical data through key pharmacokinetic parameters, allowing the assessment of the predictivity of disease models at a pharmacological level.

Finally, we present the thesis discussion – which was published separately from the general summary – in **Chapter 6**. Here, we summarise the findings from the previous chapters and put them into perspective with each other and the broader literature. We discuss the impact of the adoption of FIMD, systematic reviews and meta-analyses, and the IB-derisk tool may have on drug development, the role of each stakeholder and what the next steps are toward a sensible assessment of the use of animal models to investigate the preliminary efficacy of new drugs.

References

1. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov. 2004;3(8):711–5.

2. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. JAMA Intern Med. 2016 Dec 1;176(12):1826–33.

3. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014 Jan;32(1):40–51.

4. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can Animal Models of Disease Reliably Inform Human Studies? PLoS Med. 2010 Mar 30;7(3):e1000245.

5. Schulz JB, Cookson MR, Hausmann L. The impact of fraudulent and irreproducible data to the translational research crisis - solutions and implementation. J Neurochem. 2016;139 Suppl 2:253–70.

6. Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans? BMJ. 2004 Feb 28;328(7438):514–7.

7. Pound P, Bracken MB. Is animal research sufficiently evidence based to be a cornerstone of biomedical research? BMJ. 2014 May 30;348:g3387–g3387.

8. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov. 2011 Sep;10(9):712–712.

9. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012 Mar 28;483(7391):531–3.

10. Juni P. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ. 2001 Jul 7;323(7303):42–6.

11. Bebarta V, Luyten D, Heard K. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad Emerg Med. 2003 Jun;10(6):684–7.

12. Vogt L, Reichlin TS, Nathues C, Würbel H. Authorization of Animal Experiments Is Based on Confidence Rather than Evidence of Scientific Rigor. Dirnagl U, editor. PLoS Biol. 2016 Dec 2;14(12):e2000598.

13. Macleod MR, McLean AL, Kyriakopoulou A, Serghiou S, De Wilde A, Sherratt N, et al. Risk of bias in reports of in vivo research: a focus for improvement. Vol. 13, PLOS Biology. 2015. p. 232097 bytes.

14. Baker D, Lidster K, Sottomayor A, Amor S. Two Years Later: Journals Are Not Yet Enforcing the ARRIVE Guidelines on Reporting Standards for Pre-Clinical Animal Studies. Eisen JA, editor. PLoS Biol. 2014 Jan 7;12(1):e1001756.

15. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVEd: Support for the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. Pritchett-Corning KR, editor. PLoS ONE. 2018 May 24;13(5):e0197882.

16. Smith AJ, Clutton RE, Lilley E, Hansen KEA, Brattelid T. PREPARE: guidelines for planning animal research and testing. Lab Anim. 2018 Apr;52(2):135–41.

17. EQIPD. European Quality in Preclinical Data (EQIPD) [Internet]. [cited 2020 Mar 6]. Available from: https://quality-preclinical-data.eu/

18. Schmidt-Pogoda A, Bonberg N, Koecke MHM, Strecker J, Wellmann J, Bruckmann N, et al. Why Most Acute Stroke Studies Are Positive in Animals but Not in Patients: A Systematic Comparison of Preclinical, Early Phase, and Phase 3 Clinical Trials of Neuroprotective Agents. Ann Neurol. 2020 Jan;87(1):40–51.

19. Perrin S. Preclinical research: Make mouse studies work. Nature. 2014 Mar;507(7493):423–5.

20. Benatar M. Lost in translation: Treatment trials in the SOD1 mouse and in human ALS. Neurobiology of Disease. 2007 Apr;26(1):1–13.

21. McKinney WT. Animal Model of Depression: I. Review of Evidence: Implications for Research. Arch Gen Psychiatry. 1969 Aug 1;21(2):240.

22. Willner P. The validity of animal models of depression. Psychopharmacology. 1984 Apr;83(1):1–16.

23. Sams-Dodd F. Strategies to optimize the validity of disease models in the drug discovery process. Drug Discovery Today. 2006 Apr;11(7–8):355–63.

24. Denayer T, Stöhr T, Roy MV. Animal models in translational medicine: Validation and prediction. European Journal of Molecular & Clinical Medicine. 2014 Aug 27;2(1):5.

1) Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands;
2) Copernicus Institute of Sustainable Development, Innovation Studies, Utrecht University, Utrecht, The Netherlands;
3) Medicines Evaluation Board, Utrecht, The Netherlands.

2

Guilherme S. Ferreira[1],
Désirée H. Veening-Griffioen[1],
Wouter P. C. Boon[2],
Ellen H. M. Moors[2],
Christine C. Gispen-de Wied[3],
Huub Schellekens[1],
Peter J. K. van Meer[1, 3]
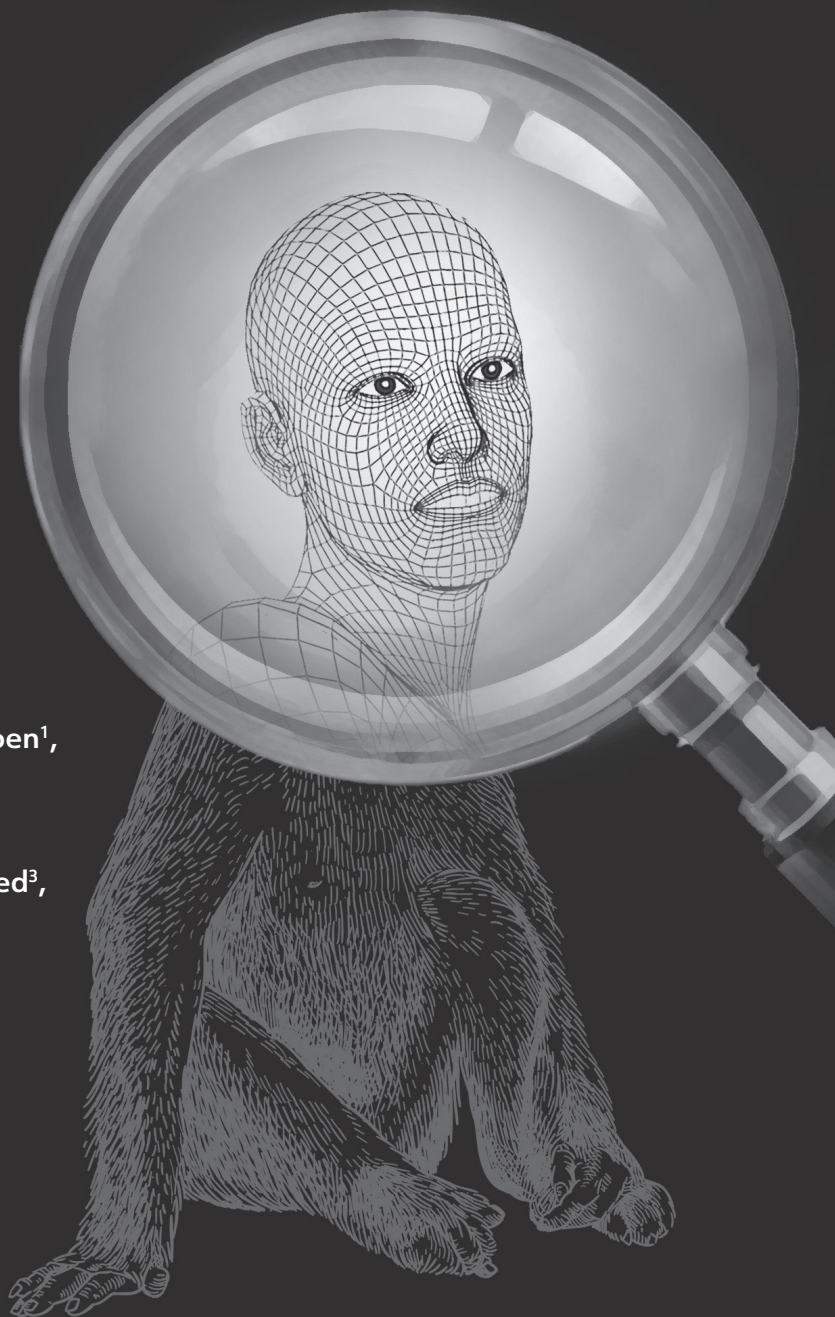
FIMD: A STANDARDISED FRAMEWORK TO VALIDATE, ASSESS AND COMPARE MODELS OF DISEASE

Abstract (Original Publication)

Introduction - Poor translation of efficacy data derived from animal models can lead to clinical trials unlikely to benefit patients – or even put them at risk – and is a potential contributor to costly and unnecessary attrition in drug development.

Objectives – To develop a tool to assess, validate and compare the clinical translatability of animal models used for the preliminary assessment of efficacy.

Design and Results – We performed a scoping review to identify the key aspects used to validate animal models. Eight domains (Epidemiology, Symptomatology and Natural History – SNH, Genetic, Biochemistry, Aetiology, Histology, Pharmacology and Endpoints) were identified. We drafted questions to evaluate the different facets of human disease simulation. We designed the Framework to Identify Models of Disease (FIMD) to include standardised instructions, a weighting and scoring system to compare models as well as factors to help interpret model similarity and evidence uncertainty. We also added a reporting quality and risk of bias assessment of drug intervention studies in the Pharmacological Validation domain. A web-based survey was conducted with experts from different stakeholders to gather input on the framework. We conducted a pilot study of the validation in two models for Type 2 Diabetes (T2D) – the ZDF rat and db/db mouse. Finally, we present a full validation and comparison of two animal models for Duchenne Muscular Dystrophy (DMD): the mdx mouse and GRMD dog. We show that there are significant differences between the mdx mouse and the GRMD dog, the latter mimicking the human epidemiological, SNH, and histological aspects to a greater extent than the mouse despite the overall lack of published data.

Conclusions - FIMD facilitates drug development by serving as the basis to select the most relevant model that can provide meaningful data and is more likely to generate translatable results to progress drug candidates to the clinic.

Keywords: animal model, drug development, efficacy model, external validity, translational research, validation, 3Rs.

Introduction

The use of non-human animals to evaluate the safety of new drugs is an integral part of the regulatory research and development process (1,2). Established at a time when laboratory animals were one of the most complex systems available, they are still considered as the gold standard. However, despite their apparent value as a drug testing system to predict safety and efficacy in humans, scientists are increasingly aware of their considerable drawbacks and limited predictivity (3–6).

While no apparent toxicity in poorly predictive animal models can lead to possible harm to patients, false toxic signals might prevent potentially safe drugs from reaching the market. At the same time, models which overestimate the efficacy of new drugs lead to clinical trials with drugs that have a modest effect at best or are entirely ineffective at worst – meaning we could be effectively putting patients at risk for no possible benefit (5,7).

We previously assessed the value of regulatory safety studies in a public-private research consortium which consisted of pharmaceutical company stakeholders, the Dutch regulatory agency and academia (8). This partnership was unique in that it allowed proprietary data to be used for our primary analyses, which could then be presented in an aggregated, anonymised fashion to propose policy changes (9–13). A key finding of these studies was that regardless of non-human primate (NHP) models having the closest biological resemblance to humans, their indiscriminate use in the safety testing of new biotechnology products (e.g. mAbs) as well as to demonstrate the similarity of biosimilar to reference products often adds limited value to the preclinical package. When taken all together, these results suggest the mandatory use of non-human animal safety testing according to current guidelines should be reconsidered.

Contrary to safety assessment, the evaluation of efficacy is not subject to formalised guidance or regulations since each new drug warrants a tailor-made approach based on its mechanism of action and indication (14,15). Consequently, predefining which assays or models to be used to test new drugs' efficacy, as done for safety, could jeopardise innovative companies' ability to develop such drug-specific strategies.

Nevertheless, most late-stage clinical trials, which are frequently based on efficacy data from non-human animal studies fail due to the lack of efficacy (16–20). The low internal validity (i.e. the methodological qualities of an experiment, such as randomisation and blinding) of non-human animal research has been frequently suggested as a likely cause for such poor translation to the clinic (21–25). For example, non-randomised and non-blinded studies are 3.4 and 3.2 times more likely to report positive results, respectively (21). Other factors, like housing and

husbandry, can also significantly affect animal models' behaviour and other phenotype parameters, such as neurotransmitter levels (26).

Several projects aimed at improving design and reporting standards of preclinical studies have been initiated in the past decade. The ARRIVE guidelines were the first harmonised guidelines to establish reporting standards in animal research (27). In the same vein, the PREPARE guidelines provide extensive guidance on the key parameters necessary to design and conduct higher quality animal experiments (28). With the growing popularity of systematic reviews and meta-analysis of non-human animal studies, the use of the GRADE approach to grade the quality of evidence and strength of the recommendations given has also been suggested (29). The European Quality In Preclinical Research (EQIPD), funded by the Innovative Medicines Initiative (IMI), was introduced in 2017 aiming at improving the robustness of preclinical data, focusing mostly on neuroscience research (30). All these initiatives now allow researchers to address these issues effectively.

The inadequate assessment of the external validity of efficacy models (i.e. how well non-human animal results are generalisable to the human situation) is also an important factor for poor translation (4). Currently, drug developers frequently rely on the well-established criteria of face, construct and predictive validity to assess the external validity of animal models (31,32). Face validity refers to how well an animal model reproduces the clinical symptoms and manifestations of a disease; construct validity refers to how close the underlying biology thought to be causing these symptoms in the human condition is replicated in the animal model; and predictive validity refers to how similar to a human an animal model responds to clinically effective drugs (32,33). Because none of these criteria is integrated nor present a systematic way to assess the ability of an animal model to predict drug efficacy in humans, they are highly susceptible to user interpretation. The absence of standardisation leads to animal models being assessed by different disease parameters, which further complicates a scientifically relevant comparison between them.

A tool applicable to *in vitro* and *in vivo* settings was suggested by Sams-Dodd and further refined by Denayer et al. (34,35). It scores five categories (species, disease simulation, face validity, complexity and predictivity) from 1 to 4, a score of 4 being the closest to the patient. Nonetheless, this tool fails to capture the nuances of most characteristics potentially relevant for the demonstration of efficacy (e.g. genes, biomarkers, histology) to make it informative and usable to evaluate the external validity of animal models.

The Framework t o Identify Models of Disease (FIMD)

Literature Review and Questionnaire Draft

To answer our research question "what are the main parameters used to validate animal models of disease?", we performed a scoping review using the search strategy "animal model"[tiab] AND validation[tiab] based on Medical Subjects Heading (MeSH) terms in PubMed on 03.06.2016. The search was not designed to be an exhaustive account of the available literature, but rather as a snapshot of the core aspects used by researchers to validate disease models. Articles were included if they were published in English and if they contained details about the validation of a new animal model or optimisation of an existing one aiming at drug screening. Articles were excluded if the aim of validation was behavioural tests or biomarkers; if the model aimed at evaluating a surgical technique; and/or if they were not available via the university's online library. Titles and abstracts were screened, and publications which met the inclusion criteria were then read in full-text.

A total of 587 records were screened by title and abstract, of which eight were not available through the university's virtual library. Eighty-one (81) articles were selected for full-text assessment. Of these, 78 records (59 of development or optimisation of animal models and 19 reviews) met the inclusion criteria. The literature search process is shown in Fig 1.
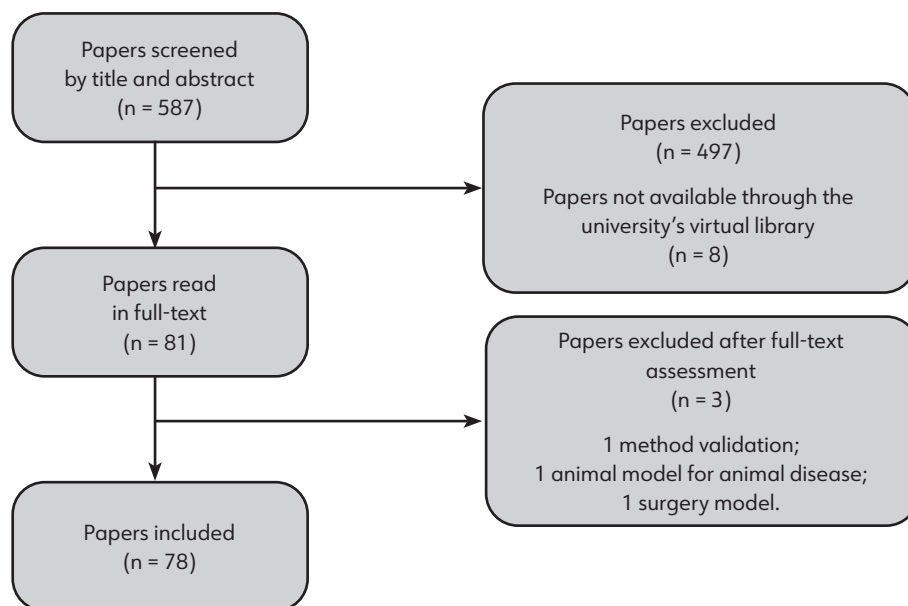

FIGURE 1 | Literature search process. Identification, screening and inclusion of articles for the identification of validation parameters in the literature.

The thematic content analysis (TCA) method was used to identify the main validation parameters present in the literature (36). Briefly, this method consists of collecting information on a field in a structured way, followed by extracting the data from the identified sources and, finally, grouping these parameters into more general, broader categories. In our study, we determined the parameters defined by the included articles' authors as determinative to 'validate' the animal model. Identified parameters were then grouped into themes – based on the concepts of face, construct and predictive validity – and were given a generic, broader name. Eight domains were identified based on the validation parameters reported in the literature and are presented in Table 1.

TABLE 1 | The eight domains identified and their frequency in the included articles.

Domain	Description	Frequency (n = 79)
Epidemiology	related to age and sex	12
Symptomatology and Natural History (SNH)	related to symptoms and natural history of disease	62
Genetics	related to genes, genetic alterations and expression	19
Biochemistry	related to biomarkers	43
Aetiology	related to the aetiology	26
Histology	related to the histopathological features	37
Pharmacological	related to the effect of effective and ineffective drugs	39
Endpoints	related to the endpoints and the methods to assess thereof chosen for the pharmacological studies	9

The eight identified domains were used to draft questions related to the different facets of disease simulation (Table 2).

We designed FIMD to circumvent some significant limitations present in the current approaches, such as the absence of standardisation, integration and facility of model comparison. We also conducted a web-based survey with experts from academia, industry and regulatory agencies to validate our findings and identify opportunities for improvement (see Appendix A). To make the framework systematic, transparent and to minimise parameter discordance, instructions on how to complete the validation sheets as well as a template are provided (see Appendixes B for instructions and online for the template[1], respectively). Non-human animal studies of both effective and ineffective drugs are included in the Pharmacological Validation. This inclusion results in a better understanding of the pathways that are involved in the disease pathophysiology of a model when compared to humans. With this information, companies with extensive non-human animal data from failed projects can easily perform a read-across of their models to inform the choice of future programmes. Also, all interventional drug studies in the Pharmacological

---

[1] Available at https://doi.org/10.1371/journal.pone.0218014.s003.

Validation have a reporting quality and risk of bias assessment adapted from the ARRIVE guidelines and the SYRCLE's Risk of Bias Tool (see Appendix B) (27,37).

TABLE 2 | The Framework to Identify Models of Disease (FIMD). Questions per validation parameter, and weight of each question and domain according to the same weighting system, in which the total of 100 points is divided equally by the eight domains, and then equally by each question within a domain.

	Weight
1. EPIDEMIOLOGICAL VALIDATION	12,5
1.1 Is the model able to simulate the disease in the relevant sexes?	6,25
1.2 Is the model able to simulate the disease in the relevant age groups (e.g. juvenile, adult or ageing)?	6,25
2. SYMPTOMATOLOGY AND NATURAL HISTORY VALIDATION	12,5
2.1 Is the model able to replicate the symptoms and co-morbidities commonly present in this disease? If so, which ones?	2,5
2.2 Is the natural history of the disease similar to human's regarding: 2.2.1 Time to onset	2,5
2.2.2 Disease progression	2,5
2.2.3 Duration of symptoms	2,5
2.2.4 Severity	2,5
3. GENETIC VALIDATION	12,5
3.1 Does this species also have orthologous genes and/or proteins involved in the human disease?	4,17
3.2 If so, are the relevant genetic mutations or alterations also present in the orthologous genes/proteins?	4,17
3.3 If so, is the expression of such orthologous genes and/or proteins similar to the human condition?	4,16
4. BIOCHEMICAL VALIDATION	12,5
4.1 If there are known pharmacodynamic (PD) biomarkers related to the pathophysiology of the disease, are they also present in the model?	3,125
4.2 Do these PD biomarkers behave similarly to humans'?	3,125
4.3 If there are known prognostic biomarkers related to the pathophysiology of the disease, are they also present in the model?	3,125
4.4 Do these prognostic biomarkers behave similarly to humans'?	3,125
5. AETIOLOGICAL VALIDATION	12,5
5.1 Is the aetiology of the disease similar to humans'?	12,5
6. HISTOLOGICAL VALIDATION	12,5
6.1 Do the histopathological structures in relevant tissues resemble the ones found in humans?	12,5
7. PHARMACOLOGICAL VALIDATION	12,5
7.1 Are effective drugs in humans also effective in this model?	4,17
7.2 Are ineffective drugs in humans also ineffective in this model?	4,17
7.3 Have drugs with different mechanisms of action and acting on different pathways been tested in this model? If so, which?	4,16
8. ENDPOINT VALIDATION	12,5
8.1 Are the endpoints used in preclinical studies the same or translatable to the clinical endpoints?	6,25
8.2 Are the methods used to assess preclinical endpoints comparable to the ones used to assess related clinical endpoints?	6,25

Weighting and Scoring System

To facilitate the comparison between animal models in the same indication and ultimately, the choice of the best fit for investigating a drug's efficacy, we developed a weighting and scoring system. This system was developed based on a maximum score of 100 points. The 100 points were equally distributed across the eight domains (12.5 points per domain) and in each domain, evenly distributed per question – the same weight (SW) system. Table 2 shows the weighting per domain and question according to the same weighting system. Although an indication-specific weighting system could potentially improve the sensitivity of the score, there is not enough data to accurately determine weightings for each question based on disease characteristics. At this stage, the presence of such a feature is hardly justifiable given the complexity it would add to the overall framework. Thus, we opted not to add indication-specific weighting systems in this version. Examples of how to score the questions are provided in Appendix B.

Calculating the score from each domain allows easy visualisation in a radar plot, which is a key product of FIMD (Box 1). The final score is an indication of the degree to which a model simulates the human condition. While it is not validated at this point, it allows researchers to identify the strengths and weaknesses of an animal model at a glance. The underlying (qualitative) data will further determine which model is the most relevant to test a given drug. Hence, the best fit will not necessarily be the model with the highest score but rather the model that more closely mimics the pathways involved in the mechanism of action of a drug.

Next to the weighting and scoring system, two factors were created to contextualise the final score: the uncertainty and the similarity factors. Although these factors are not used in the score calculation, they help to explain the differences or absence thereof in the score between models. The uncertainty factor is used to inform how much information is missing in the validation of a model – how many questions were answered with 'unclear'. 'Unclear' is used to answer questions for which no data is available or conflicting evidence is found. The similarity score is used when comparing two models, informing about how often questions received the same answer in both models. It indicates how comparable two models are. Further details on how to calculate and interpret the uncertainty and similarity factor are available in Appendix B, sections D and E.

BOX 1 | Validation Domains Overview with a Radar Plot.

The radar plot is generated by using the ratio of each domain's score (sum of each question within a domain) to the maximum possible score. For instance, if a model scores 6.25 points in the first domain, Epidemiological Validation (3.125 per question), of a total of 12.5 points, the ratio is 0.5 (6.25/12.5). These ratios are calculated for each domain and then aggregated in the radar plot, which is a relative demonstration of how close an animal model can simulate the human condition. The closer a domain is to the edge, the closer the model mimics that specific aspect of the human disease (ratio of 1). Radar plots can include one or more models. An example is given in Fig 2.
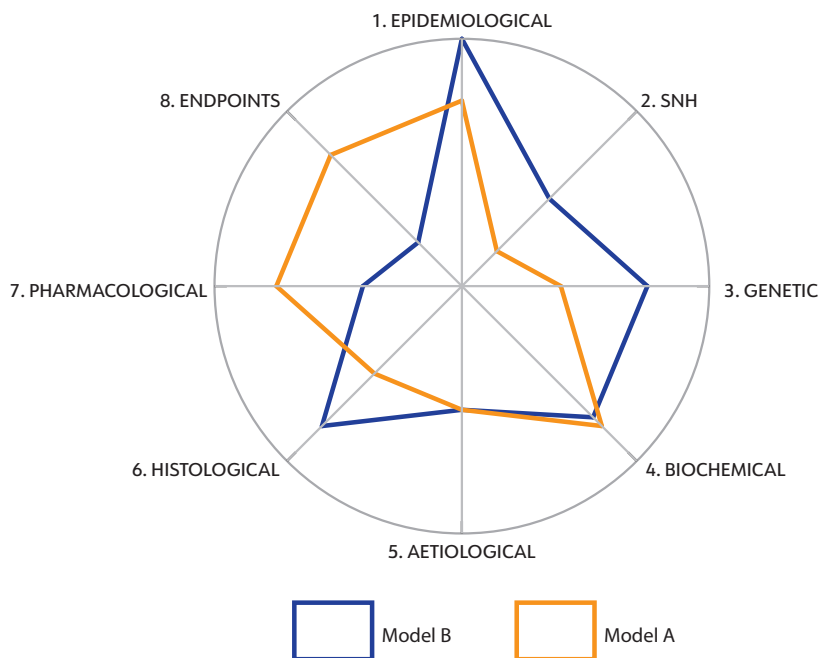
FIGURE 2 | Hypothetical Radar Plot. Radar plot with hypothetical scores per parameter per model. The closer a parameter is to the edge, the better the model simulates that aspect of the human disease. SNH – Symptomatology and Natural History.

Validating Efficacy Models with FIMD

We addressed the lack of a clear definition of what the minimum requirements to validate an animal model are from previous approaches. There are several interpretations of what the validation of an assay or model should entail and how to assure its reproducibility. For the OECD and EURL-ECVAM, validation means establishing a statistically rigid range for several criteria to describe a test's ability to be reproduced reliably (38). This validation is frequently an expensive and lengthy process, mostly applicable to *in vitro* tests, which would not be practical or add much value to efficacy assessment in animal models (14). A more applied approach is presented by ICHS5(R3), in which the definition of the context of use (i.e. the

conditions in which the assay results can be relied upon) is the guiding factor for the validation of a test (39).

Our validation definition aims to provide the evidence for an efficacy model's context of use (i.e. the intended indication for which drugs are being developed). Although animal models are not expected to mimic the human condition completely, identifying which aspects they can reproduce and to which extent is crucial to determining whether they are the most suitable choice. Therefore, we established four levels of confidence in the validation of animal models in their context of use based on the percentage of definite answers to the eight validation domains (see Table 3). A 'definite answer' is defined as any answer except for 'unclear'.

TABLE 3 | Different levels of validation according to the percentage of definite answers to the questionnaire.

(%) Definite Answers	Validation Level
0 - 40	Insufficiently validated
41 - 60	Slightly validated
61 - 80	Moderately validated
81 - 100	Highly validated

The product of FIMD is a validation sheet of an animal model for an indication, which provides the necessary information for its assessment as a potential model to demonstrate a drug's efficacy. We have conducted a pilot study and a full validation in one indication each to test the main features of FIMD, which are further described in **Chapter 3**.

Limitations

In FIMD, the validation of an animal model relies on the definition of the disease parameters and therefore, it depends on how well-understood the aetiology and pathophysiology are. Poorly understood diseases (e.g. Alzheimer's) represent a major challenge for the definition of these parameters and thus to the validation of animal models in these indications. However, FIMD provides researchers with a platform to discuss both the more established aspects of a disease as well as the ones in which there is no consensus in a structured and standardised manner (e.g. aetiology, genetic, biochemical). This way, an animal model's ability to simulate the different aspects of disease according to various theories or hypotheses (whenever available) can be discussed and assessed in one place by the same parameters.

All in all, FIMD provides researchers with the information necessary for them to decide on whether a model is the most relevant to test the efficacy of a new drug. Additionally, the definition of the disease parameters themselves might prove controversial, as there is currently no standardised and objective way of defining

them. The recommendation to organise a task force with experts in the field to achieve consensus on which parameters to include could preclude the introduction of most biases, although it will frequently be unfeasible.

The weighting system provided here aims to roughly indicate in quantitative terms how well a model can simulate the human disease, but it could not be validated yet. Ideally, the weight of each question should be set based on a statistical model that correlates the likelihood of an animal model predicting human response with its assessment through FIMD. This approach would first require that a considerable number of models are assessed through FIMD; followed by an evaluation of the likelihood each model has of predicting efficacy in humans for each drug that has been tested on it. Then, it would be possible to correlate the likelihood of a model predicting human response with each type of answer to the validation questions. This statistical model would give a weighting for each question and also allow the setting of the weights specific to each indication, which would provide a more accurate and scientifically sound weighting system. Such an approach will become feasible as more animal models are validated using FIMD. Furthermore, the use of partial grades (50% of the score) for models which only partially mimic a particular aspect of the disease considerably decreases sensitivity, leading to the same score for models that simulate the same aspect to different extents.

Another limitation is that FIMD only includes publicly available information, which means it inherently incorporates the publication bias often reported in preclinical and clinical research (4,40,41). However, there is a growing demand for the pre-registration of preclinical studies and the publication of their results (42–44). With the establishment of registries like PreclinicalTrials.eu, the overall publication bias is expected to be reduced and therefore, so will be its impact on FIMD (45). This factor is especially relevant for the Pharmacological Validation questions 7.1 and 7.2, which include drug intervention studies and might lead to skewed results. Also, due to time and resource constraints, we did not investigate the inter-rater reliability to verify the reproducibility of the results. However, we provide detailed instructions in Appendix B which are likely to reduce variation between users.

Applications and Final Considerations

Based on a drug's mechanism of action, models can be first discriminated by assessing whether the correlation between non-human animal and human drug studies of relevant pathways is available in the Pharmacological Validation domain. The other domains provide additional information, such as the presence of relevant genes and biomarkers. Finally, the validation level is an index of the reliability of a model's overall ability to mimic the human condition, serving as another layer to further differentiate potentially more useful from less useful models. The

FIMD: a standardised framework to validate, assess and compare models of disease

combination of all these features allows researchers to select, among a plethora of models, the model most likely to correctly predict the efficacy of a drug in humans.

FIMD allows the comparison between animal models from a purely scientific perspective. In practice, other factors may influence the choice of one model over another, such as personnel expertise, available facilities and cost. Larger models (e.g. dogs) are considerably more expensive than rodents and, despite possibly mimicking the human disease to a greater extent, might still not be the model of choice in many occasions. These factors are not taken into account in FIMD, as they vary not only between but also within institutions. However, by providing the scientific background on which to base the selection of a model, we hope to substantiate discussions within ethics committees and funding bodies on how these non-scientific factors can affect model selection. This discussion can provide researchers with arguments to obtain more funding to conduct projects on more expensive models which also have a higher likelihood of predicting the efficacy of new drugs.

An essential application for FIMD is on the approval of non-human animal studies by Institutional Review Boards (IRBs). IRBs often base their decisions on unpublished non-human animal studies with low internal validity (44). FIMD presents an opportunity to assess the validity of a specific model for efficacy assessment while also providing insights from earlier research. By using the validation sheet of models used to support the first-in-human trials, IRBs can, for the first time, tackle all these issues at once. FIMD provides the background for the choice of the model(s), allowing IRBs to accurately assess whether the data generated is likely to be translatable to the clinic.

Although it is not our goal to address the problems with internal validity of animal studies with FIMD, we included a reporting quality and risk of bias assessment in the Pharmacological Validation. We aimed to confront researchers with the often-poor quality of studies routinely used to base grant and marketing authorisation decisions, and subsequent research, forcing them to think about their own study designs. For this reason, we did not provide guidance on how to interpret these results as the relevance of each quality aspect will depend on the research question and indication.

The collection of validation sheets of models of efficacy will allow the development of an open database of validated models in which users can, based on their drug's characteristics (e.g. mechanism of action or intended indication), find the optimal model to evaluate the efficacy of a drug before planning an animal experiment.

FIMD can simplify the interaction between companies and regulatory agencies as

it permits a more objective and science-based discussion on the selection of an animal model. It can potentially prevent efficacy studies on less relevant models, effectively contributing to the reduction of the use of animal models in the context of the 3R's. By assessing the degree to which a model mimics a human condition, FIMD facilitates the choice of an appropriate model for efficacy assessment and promotes the conduct of efficacy studies whose results are more likely to translate to the clinical situation.

Acknowledgements

Competing interests

References

1. European Commission. DIRECTIVE 2004/27/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 31 March 2004 amending Directive 2001/83/EC on the Community code relating to medicinal products for human use [Internet]. [cited 2018 Jun 13]. Available from: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/dir_2004_27/dir_2004_27_en.pdf

2. US Congress. Federal Food, Drug, and Cosmetic Act (FD&C Act) [Internet]. [cited 2018 Jun 13]. Available from: https://www.fda.gov/regulatory-information/laws-enforced-fda/federal-food-drug-and-cosmetic-act-fdc-act

3. Hackam DG, Redelmeier DA. Translation of Research Evidence From Animals to Humans. JAMA. 2006 Oct 11;296(14):1727.

4. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can Animal Models of Disease Reliably Inform Human Studies? PLoS Medicine. 2010 Mar 30;7(3):1–8.

5. Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans? BMJ : British Medical Journal. 2004;328(7438):514–7.

6. Godlee F. We need better animal research, better reported. BMJ. 2018 Jan 11;k124.

7. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ : British Medical Journal. 2007;334(7586):197–197.

8. Top Institute Pharma. Predictive Value of Animal Testing Research [Internet]. [cited 2018 May 7]. Available from: http://www.tipharma.com/pharmaceutical-research-projects/completed-projects/predictive-value-of-animal-testing/

9. van Meer PJK, Kooijman M, Gispen-de Wied CC, Moors EHM, Schellekens H. The ability of animal studies to detect serious post marketing adverse events is limited. Regulatory Toxicology and Pharmacology. 2012;64(3):345–9.

10. van Meer PJK, Kooijman M, van der Laan JW, Moors EHM, Schellekens H. The value of non-human primates in the development of monoclonal antibodies. Nature Biotechnology. 2013 Oct 1;31(10):882–3.

11. Kooijman M, van Meer PJK, Gispen-de Wied CC, Moors EHM, Hekkert MP, Schellekens H. The risk-based approach to ATMP development - Generally accepted by regulators but infrequently used by companies. Regulatory Toxicology and Pharmacology. 2013;67(2):221–5.

12. Kooijman M, Hekkert MP, van Meer PJK, Moors EHM, Schellekens H. How institutional logics hamper innovation: The case of animal testing. Technological Forecasting and Social Change. 2017;118:70–9.

13. van Meer PJK, Ebbers HC, Kooijman M, Gispen-de Wied CC, Silva-Lima B, Moors EHM, et al. Contribution of animal studies to evaluate the similarity of biosimilars to refrence products. Drug Discovery Today. 2015;20(4):483–90.

14. Varga OE, Hansen AK, Sandoe P, Olsson AS. Validating Animal Models for Preclinical Research: A Scientific and Ethical Discussion. Atla. 2010;38(3):245–8.

15. Langhof H, Chin WWL, Wieschowski S, Federico C, Kimmelman J, Strech D. Preclinical efficacy in therapeutic area guidelines from the U.S. Food and Drug Administration and the European Medicines Agency: a cross-sectional study. British Journal of Pharmacology. 2018;175(22):4229–38.

16. Arrowsmith J. Trial watch: phase III and submission failures: 2007-2010. Nature reviews Drug discovery. 2011;10(2):87.

17. Arrowsmith J, Miller P. Trial Watch: Phase II and Phase III attrition rates 2011-2012. Nature reviews Drug discovery. 2013;12(8):569.

18. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotech. 2014;32(1):40–51.

19. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. JAMA Internal Medicine. 2016;176(12):1826.

20. Schulz JB, Cookson MR, Hausmann L. The impact of fraudulent and irreproducible data to the translational research crisis – solutions and implementation. Journal of Neurochemistry. 2016;139:253–70.

21. Bebarta V, Luyten D, Heard K. Emergency medicine animal research: Does use of randomization and blinding affect the results? Academic Emergency Medicine. 2003;10(6):684–7.

22. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature. 2012;490(7419):187–91.

23. Peers IS, South MC, Ceuppens PR, Bright JD, Pilling E. Can you trust your animal study data? Nature Reviews Drug Discovery. 2014;13(7):560.

24. Prinz F, Schlange T, Asadullah K. Believe it or not: How much can we rely on published data on potential drug targets? Nature Reviews Drug Discovery. 2011;10(9):712–3.

25. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012;483(7391):531–3.

26. Pasquarelli N, Voehringer P, Henke J, Ferger B. Effect of a change in housing conditions on body weight, behavior and brain neurotransmitters in male C57BL/6J mice. Behavioural Brain Research. 2017;333:35–42.

27. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLOS Biology. 2010 Jun 29;8(6):e1000412.

28. Smith AJ, Clutton RE, Lilley E, Hansen KEA, Brattelid T. PREPARE: guidelines for planning animal research and testing. Laboratory Animals. 2017 Aug 3;1–7.

29. Wei D, Tang K, Wang Q, Estill J, Yao L, Wang X, et al. The use of GRADE approach in systematic reviews of animal studies. Journal of Evidence-Based Medicine. 2016;9(2):98–104.

30. EQIPD. European Quality In Preclinical Data [Internet]. [cited 2018 Oct 4]. Available from: https://quality-preclinical-data.eu/

31. McKinney Jr. WT, Bunney Jr. WE. Animal model of depression. I. Review of evidence: implications for research. ArchGenPsychiatry. 1969;21(2):240–8.

32. Willner P. The validity of animal models of depression. Psychopharmacology. 1984;83(1):1–16.

33. McGonigle P, Ruggeri B. Animal models of human disease: Challenges in enabling translation. Biochemical Pharmacology. 2014;87(1):162–71.

34. Sams-Dodd F. Strategies to optimize the validity of disease models in the drug discovery process. Drug Discovery Today. 2006;11(7–8):355–63.

35. Denayer T, Stöhrn T, Van Roy M. Animal models in translational medicine: Validation and prediction. New Horizons in Translational Medicine. 2014;2(1):5–11.

36. Elo S, Kyngäs H. The qualitative content analysis process. Journal of Advanced Nursing. 2008;62(1):107–15.

37. Hooijmans CR, Rovers MM, Vries RBM De, Leenaars M, Ritskes-hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC Medical Research Methodology. 2014;14(1):1–9.

38. OECD. Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment [Internet]. Vol. 34. 2005 [cited 2016 Jul 4]. p. 1–96. Available from: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2005)14

39. ICH. S5(R3) Revision of S5 Guideline on Detection of Toxicity to Reproduction for Human Pharmaceuticals [Internet]. 2017 [cited 2018 Mar 17]. p. 1–63. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S5/S5-R3EWG_Step2_Guideline_2017_0705.pdf

40. Dwan K, Altman DG, Arnaiz J a., Bloom J, Chan A-W, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS Biology. 2008;3(8):e3081.

FIMD: a standardised framework to validate, assess and compare models of disease

41. Sena ES, Bart van der Worp H, Bath PMW, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biology. 2010;8(3):e1000344.
42. Kimmelman J, Anderson JA. Should preclinical studies be registered? Nature Biotechnology. 2012;30(6):488–9.
43. Jansen of Lorkeers SJ, Doevendans PA, Chamuleau SAJ. All preclinical trials should be registered in advance in an online registry. European journal of clinical investigation. 2014;44(9):891–2.
44. Wieschowski S, Chin WWL, Federico C, Sievers S, Kimmelman J, Strech D. Preclinical efficacy studies in investigator brochures: Do they enable risk–benefit assessment? PLOS Biology. 2018;16(4):e2004879.
45. PreclinicalTrials.eu. About PreclinicalTrials.eu [Internet]. [cited 2018 Apr 14]. Available from: https://preclinicaltrials.eu/#aboutSitePage

Appendix A – Survey

Survey Results

A web-based survey was created to collect feedback regarding the use and implementation of FIMD (formerly Question-Based Validation Sheet - QBVS) from experts from different stakeholders (e.g. regulatory agencies, academia and pharmaceutical industry). The expert panel had access to the questionnaire, instructions on how to complete it and the results of the case studies. Their suggestions were processed, considered relevant and therefore, implemented in FIMD.

The anonymous survey could be accessed via an open link, which means any person with the link could access it. The link was initially sent by email to 46 experts, which were identified by convenience sample (i.e. use of authors' network, contacting corresponding authors of studies included in the pilot study). The distribution of the survey link within the experts' network was not only allowed but also encouraged. Nonetheless, due to the low number of responses, only descriptive statistics are presented.

Twelve (12) experts participated in the survey. The majority is based in Europe (83%), are academics (67%), but industry and regulatory agencies were also represented (25% and 8%, respectively). All experts had experience with designing animal studies, 92% have conducted animal experiments, 83% have assessed and reviewed animal studies, 67% have developed or optimised animal models and 25% have experience on evaluating the risk of bias and publication bias, drafting guidelines for safety aspects or regulatory aspects in general. Most respondents (75%) have at least ten years of experience in these functions, 17% have at least five and 8% between two and five years.

Most respondents support the use of FIMD as a tool to validate animal models of disease (92%). Around 83% responded it is only moderately likely for two independent researchers to arrive at the same results, despite the availability of instructions on how to fill in and score each question. The main reason reported was the difference in background and the individual skill of each researcher.

Regarding adoption, almost 42% of those surveyed answered that is 'slightly likely' for FIMD to be adopted in their organisation, followed by 33% answering it is 'very likely' while 27% thinks it is only 'moderately likely'. The main reasons reported are the significant time investment for training and execution of the validation.

As for the score, most experts think it adds value to the validation process (83%), although 17% sees it also as a potential source of bias. More than half affirmed that

the score should be defined only in the method (58%), not allowing researchers to set the weights themselves.

One respondent commented that, for some animal models, there are significant differences in pathophysiological parameters between colonies, which should be reported. One example is the GRMD dog, which has a different natural history of the disease in some colonies: in one American colony, dogs hardly ever lose ambulation while in the French colony, around 25% lose the ability to walk (1). If the phenotype differences are substantial enough, they shall be investigated beforehand, and, in such cases, each colony would be considered as a related, but an individual model. In this situation, only studies which included animals from that specific colony shall be included. These differences shall also be reported in the 'Background Information' section of FIMD for all related models.

Another respondent suggested that a parameter should be included to indicate how many articles support the answers in each section of FIMD. This parameter was called 'strength of evidence': the more well-established the evidence becomes, the higher the strength of evidence. We decided not to include it because to make it useful; we would have to apply it to all questions, unreasonably increasing the complexity of the framework and hampering its feasibility.

An updated to the weighting system was also suggested by the expert panel. Where FIMD initially supported a one-size-fits-all approach to defining the weighting, the development of a fit-for-purpose weighting system was recommended to improve the adoption of FIMD, especially in the industry. Such a strategy aimed at increasing the resolution of the scoring system. Nonetheless, since there are not enough data to determine the values accurately, we have not been able to test whether the different weighting systems, in fact, increase resolution. At this stage, the presence of such a feature is hardly justifiable, given the complexity it adds to the overall framework. Thus, we decided not to include the indication-specific weighting. We have maintained only the same weighting system, in which all domains (and questions within a domain) have the same weighting.

Web-Survey Questions

Q0 Which of the examples below is clearer in helping to understand how to answer the following question: 2.1.2 Is the model able to simulate the disease in the relevant age groups (juvenile, adult or ageing)?
1. Specific Example
Yes, completely.
The ZDF rat develops diabetes around the same age as humans – middle-aged adults, progressing into old age and death.

Yes, partially.
Although the time to onset of diabetes has been decreasing in the past years, ZDF rat still develops diabetes somewhat earlier than humans. However, it progresses into adulthood and ageing phases, similarly to the human disease.

No.
The ZDF rat develops diabetes much earlier than humans already in their first two weeks. The disease progresses and leads to death before the beginning of adulthood.

2. Generic Example
Yes, completely.
The model X develops disease Y around the same age as humans – middle-aged adults, progressing into old age and death.

Yes, partially.
Although the time to onset of disease Y has been decreasing in the past years, the model X still develops disease Y somewhat earlier than humans. However, it progresses into adulthood and ageing phases, similarly to the human disease.

No.
The model X develops disease Y much earlier than humans already in their first two weeks. The disease progresses and leads to death before the beginning of adulthood.

Q11 Based on the instructions provided and your own experience, would you be able to fill out the QBVS for a given animal model?
1. Yes
2. No
3. I am not sure

Q111 What would prevent you from being able to fill out the QBVS?
1. Questions not clear enough
2. Instructions not clear enough
3. Instructions not objective enough
4. Other

Q12 If different people fill out the QBVS for the same model, how likely is it that they would arrive at the same results?
1. Not at all likely
2. Slightly likely
3. Moderately likely
4. Very likely
5. Completely likely

Q121 Why do you think they would not arrive at the same results?
Open text.

Q13 Do you think the QBVS can be used as a tool to validate animal models?
1. Yes
2. No

Q131 Why do you think the QBVS cannot be used as a tool to validate animal models?
Open text.

Q14 Would you include, edit or exclude any questions?
1. Yes
2. No

Q141 Which changes would you suggest?
Open text.

Q21 Do you think having a score included in the QBVS...
1. adds value to it?
2. introduces bias?
3. Other

Q211 How would you determine the weight for each question?
Open text.

Q212 For what reason(s) would it not add value?
Open text.

Q213 For what reason(s) would it introduce bias?
Open text.

Q22 Having a score included in the QBVS would affect my opinion of it...
1. negatively
2. positively
3. neither negatively nor positively

Q23 If the QBVS were to have a score included, do you think the weighting system(s)...
1. should be defined only in the QBVS method
2. should be defined only by the person filling out the QBVS
3. should be defined in the QBVS method and also allow the person filling out the QBVS to set its own
4. I do not think QBVS should have a score embedded
5. none of the above

Q31 From the presented weighting methods, which represents the best way to weight different aspects of disease simulation in animal models?
1. SW
2. W1
3. W3
4. None of the above.
5. I do not know.

Q32 Would you make any changes to the weights of specific questions?
Open text.

Q41 What is the likelihood of the QBVS tool being adopted by you and your colleagues?
1. Not at all likely
2. Slightly likely
3. Moderately likely
4. Very likely
5. Completely likely

Q411 Why is it unlikely you and your colleagues would adopt the QBVS?
1. the time investment is too big
2. the instructions are not clear enough
3. the method is not objective enough
4. too much training required
5. other

Q412 What changes should be made for you to adopt the QBVS?
Open text.

Q51 Do you have any additional remarks not covered by any of the previous questions?
Open text.

Q52 Would you know anyone who could be also interested in participating in this consultation? If so, please leave their name and email for contact below.
Open text.

Q61 In which region are you located?
1. USA and Canada
2. Latin America
3. Europe
4. Africa
5. Middle East
6. Asia
7. Australia

Q62 What is your position at your institution?
Open text.

Q63 For what kind of institution do you work?
1. Government Body
2. Industry (1-10 employees)
3. Industry (11-50 employees)
4. Industry (51-250 employees)
5. Industry (251+ employees)
6. Regulatory Agency
7. Research Institute (not within a University)
8. University
9. Other

Q64 In which activities presented below are or have you been involved? (mark all that apply)
1. Design of animal studies
2. Conduct of animal studies
3. Animal model development and/or optimisation
4. Assessment and review of animal studies (including due diligence)
5. Other

Q65 How many years of experience do you have in these functions?
1. Up to 2 years
2. 2+ to 5 years
3. 5+ to 10 years
4. 10+ years.

Appendix B - Instructions

A. Getting Started

The definition of the validation parameters is the first step to validate an animal model in a specific indication. While we recommend a task force of experts in the field as the most objective and reliable alternative to define these parameters, we realise the difficulties to set one up will often prevent its implementation. In these cases, the disease parameters can be identified by searching for published reviews on the disease and its natural history. Only full-text articles in English and published in peer-reviewed journals shall be included.

To identify the correct terminology for both the indication and model of interest, we suggest the use of the Medical Subject Headings (MeSH) website[2] for PubMed; and the Emtree[3] for Embase. For animal models, it is important to include abbreviations and plurals as well (*"ZDF rat" OR "ZDF rats" OR "Zucker Diabetic Fatty rat" OR "Zucker Diabetic Fatty rats"*). Besides reviews published in the literature, typing the strain of the model (e.g. BKS.Cg-Dock7m +/+ Leprdb/J for the db/db mouse) on the supplier's website (e.g. Jackson Laboratories, Charles River Laboratories) often provides extra information on the basic characterisation of the strain. Also, whenever different background strains can be used to breed the model, it shall be investigated whether there are significant differences between different strains. If so, separate validation sheets must be filled for each background strain of interest. If the background strain is not mentioned, it shall be reported as such.

Once relevant reviews are identified, the reference list and forward citation (e.g. Google Scholar, Web of Science, Scopus) shall be used to determine original research to answer the validation questions. Since completing the questionnaire is mostly an iterative process, no specific search string is needed to answer most questions, except for the Pharmacological Validation questions 7.1 and 7.2, for which further guidance is provided in their respective sections. In the case that after the general literature search some questions are still left unanswered, specific search strings can be used to identify relevant articles. For example, if no information was initially found about the up- or downregulation of a gene in a model, PubMed and Embase can be searched with a string such as '[model, including terminology variations] AND [gene]'. The use of MeSH and Emtree terms for each search component is encouraged.

---

[2]   https://meshb.nlm.nih.gov/search

[3]   https://www.embase.com/#emtreeSearch/default

The final score is based on the Same Weighting (SW) system. Each domain has the same weight in a total of 100 points (12.5 per domain). The 12.5 points are equally divided by the number of questions in each domain. Based on the answers inputted in the Score Calculator (available online as S8 Supporting Information), a radar plot will be automatically generated.

Once the final score, the uncertainty and similarity factors are calculated (sections D and E, respectively), fill in this information in the first page of the validation sheet. Include the validation date (latest date of publication of the validation references). Determine the validation level as defined in Table 3 from the main text. Describe the historical background of the model as described in section B. Fill each question's score.

Below we present a quick stepwise approach to FIMD:

1) Use MeSH and Emtree to identify all terminology related to the model and indication of interest, including abbreviations and plurals;
2) Organise an expert group to define the disease parameters. If that is not possible, search for reviews on the human disease and animal models for that condition. Whenever necessary, justify the choice of disease parameters;
3) Answer the questions in the validation sheet template (available online as S3 Supporting Information) based on the information found in the papers. If, in the end, some questions are still unanswered, use specific search strings for these questions;
4) For the Pharmacological Validation, assess each paper based on the instructions for questions 7.1, 7.2; and sections F1 Reporting Quality and F2 Risk of Bias. Fill this information in a separate spreadsheet file (available online as S7 Supporting Information);
5) Fill in the Score Calculator (available online as S8 Supporting Information) to get the total score and the radar plot;
6) Calculate the Uncertainty Factor according to section D;
7) If another model has been validated for the same indication, calculate the Similarity Factor according to section E;
8) Finalise the validation sheet (fill in the first page validation sheet, score per question).

B. General Definitions

<u>Answers</u>
All questions (except for 7.1 and 7.2) shall be answered with 'yes', 'yes, completely', 'yes, partially', 'no' or 'unclear'. Examples based on the results of the pilot study of Type 2 Diabetes (T2D) and the full validation of Duchenne Muscular Dystrophy (DMD) models are provided in the *How to score?* grey-lighted boxes to illustrate how

the questions should be answered. Given the currently limited available dataset, it was not possible to provide data-driven examples for all possible answers to all questions. For each question, one answer was extracted as it is from Appendices D and E and hypothetical answers were provided for all the other possible answers.

'Yes' and 'Yes, completely' have the same meaning and are equivalent. The difference in use depends on whether the question allows a gradation of response. For example, question 3.1 refers to the presence of orthologous genes and proteins in the model, for which the answer can only be 'yes', 'unclear' or 'no' since each gene is assessed individually. On the other hand, answering question 5.1 with a 'yes' is not enough as the question allows the gradation: animal models may only partially mimic the human condition's aetiology. Questions answered with 'Yes' or 'Yes, completely' get 100% of the score.

'Yes, partially' is used to answer questions in which the model only recapitulates parts of the human disease parameter. For instance, if for question 1.1, a human condition can manifest in both males and females, but the animal model can only simulate the disease in males, the answer shall be 'yes, partially' as the model can only partially mimic that aspect. Questions answered with 'Yes, partially' get 50% of the maximum score.

'No' shall be used to answer questions when the model does not mimic that aspect of the human condition at all. For one question (1.1) we did not include a 'no' answer in the examples because such an answer would require an animal model to simulate the human condition in the opposite sex the condition manifest in humans, an unlikely situation. Questions answered with 'No' get 0% of the maximum score.

'Unclear' shall be used in two situations. The first is when no literature can be found even after the use of specific search strings. The second is to answer questions for which conflicting evidence is found. For example, if when answering question 3.3 two papers are found: one stating there is no change in a gene's regulation while another shows upregulation of said gene, this characterises conflicting evidence. A brief discussion on the results available is warranted, which shall also include an assessment of whether one or another is more likely to be true if possible. Special attention is necessary in cases of old vs new studies, in which new evidence might be brought to light which contradicts previous research. Whenever possible, these differences must be investigated to evaluate whether there is a new consensus in the field that previous data is not considered accurate any longer. We encourage researchers to always search for recent papers to further corroborate old research findings (e.g. up to the 2000s). Questions answered with 'Unclear' get 10% of the maximum score.

Well-established
The term 'well-established' is used in some questions to indicate a scientific consensus on a topic. Its subjectivity is intentional, and it requires a scientifically sound and well-referenced justification.

C. Fields Breakdown

Model Name
The name of the model shall be presented with as much detail as possible, including specific genetic alterations whenever relevant (e.g. Zucker Diabetic Fatty (ZDF)-$Lepr^{fa}$/Crl rat). In case a model has a much commoner name, this shall be stated in parenthesis (e.g. C57Bl10scsn-Dmdmdx mouse (mdx mouse)).

Indication
The indication for which the model is being validated shall be stated. In case of symptoms of diseases, the disease in which the chosen symptom is manifested shall also be stated (e.g. food poisoning diarrhoea).

Validation Date
The validation date refers to the publication date of the last peer-reviewed reference included in any of the eight domains.

Subsections
Subsections can either be questions for which no further distinction is necessary (e.g. 5.1 on aetiology) or each section within a question (e.g. in question 7.1 on effective drugs, each drug is considered a subsection). For instance, a specific indication had its parameters defined as 5 symptoms, 4 genes, 2 pharmacodynamic and 2 prognostic biomarkers, 4 histopathological features, 1 approved drug class with 2 drugs approved and 2 failed drug classes with one drug each. The total number of subsections would be: 2 (epidemiological) + 9 (SNH) + 12 (genetic) + 8 (biochemical) + 1 (aetiological) + 4 (histological) + 5 (pharmacological) + 2 (endpoints), which is equal to 43. This number is the denominator used to calculate the uncertainty and similarity factors (sections D and E, respectively).

Total Score
The total score is the sum of each question score for all eight domains. It is calculated automatically by the Score Calculator (available online as S8 Supporting Information).

Validation Level
The validation level refers to the level of confidence in the reliability of a model to simulate the human condition, and it is defined based on the percentage of definite answers in the validation sheet. A 'definite answer' is defined as any answer except

for 'unclear', which is used to indicate the absence of evidence in the literature or conflicting results. The validation level can be insufficiently validated (up to 40% of definite answers); slightly validated (41-60%); moderately validated (61-80%); and highly validated (81-100%).

Historical Background

A brief historical background of the model shall be presented. It shall include the first publication to mention the model whenever possible as well as a brief history line of its characterisation. Known colony and supplier variations shall be stated at the end with the relevant references.

1. Epidemiological Validation

1.1 Is the model able to simulate the disease in the relevant sexes?

Describe whether the model can reproduce the disease in both sexes with sex-specific characteristics in a similar fashion to how the disease is manifested in humans. Sex-specific information on the pathophysiology of the disease shall be included if present (e.g. protective effect of oestrogen in females for the development of diabetes).

How to score?

Yes, completely.
Both male and female db/db mice can develop diabetes. The diabetic condition is somewhat more pronounced in male than in female mice. This specific female resistance to the development of diabetes is in line with what is seen in humans, due to a possible protective effect of oestrogen on pancreatic beta-cells.

Yes, partially.
Only male db/db mice can develop diabetes. Female mice have slightly higher but non-significant glycaemia and do not develop overt diabetes at any point in life.

1.2 Is the model able to simulate the disease in the relevant age groups (e.g. juvenile, adult or ageing)?

Describe whether the model can reproduce (spontaneously or by intervention) the disease in the relevant age groups (e.g. juvenile, adult and ageing) in which the disease is commonly manifested in humans (clinical presentation of symptoms). It is important to state in which age group is the onset of the disease in the model when compared to humans and whether the disease develops totally, partially or not at all in the same development phases. Where possible, narrower age categories shall be reported.

How to score?

Yes, completely.
The ZDF rat develops diabetes around the same age as humans – middle-aged adults, progressing into old age and leading to death.

Yes, partially.
Although the time to onset of diabetes has been decreasing in the past years, the ZDF rat still develops diabetes somewhat earlier than humans. However, it progresses into adulthood and ageing phases, similarly to the human disease.

No.
The ZDF rat develops diabetes much earlier than humans already in their first two weeks. The disease progresses and leads to death before the beginning of adulthood.

2. Symptomatology and Natural History (SNH) Validation

2.1 Is the model able to replicate the symptoms and co-morbidities commonly present in this disease? If so, which ones?

Describe whether the hallmark symptoms and co-morbidities of the disease are present and how well (completely, partially or not at all) the model can simulate these symptoms/co-morbidities. The definition of which symptoms/co-morbidities to include shall ideally be done by a working group of experts in the field. Alternatively, a search on scientific literature databases such as PubMed and Embase can be performed to identify reviews on the human condition from which the characteristic symptoms and co-morbidities can be identified. A brief description of each symptom in the model, when compared to the human condition, shall be added. Sometimes, human symptoms cannot be directly translated to the animal situation or are a result of a more fundamental process that can be compared in this section instead of the nominal symptoms. For instance, in T2D, the symptoms presented by humans (e.g. polyphagia, polydipsia, polyuria, weight loss) are directly caused by hyperglycaemia, hyper and hypoinsulinemia, and dyslipidaemia. We considered the presence of the latter three together with obesity – which is an important co-morbidity – as more relevant parameters to validate the animal models for T2D. In such cases, the justification must be presented in this section before the comparison between animal and human parameters.

> **How to score?**
>
> The total score for this question is calculated based on the proportionality of symptoms modelled and partially modelled. For example, if there are three symptoms defined for a given disease and a model simulates one symptom completely (whole point), one partially (half a point) and the last one not at all, the score would be calculated by multiplying 1.5/3 by the weight of this question.
>
> **Examples**
>
> *Symptoms modelled.*
> Muscle wasting: comparable to DMD boys across all main groups of muscles.
>
> *Symptoms partially modelled.*
> Muscle wasting: comparable to DMD boys only regarding diaphragm degeneration.
>
> *Symptoms not modelled.*
> Muscle wasting: the mdx mouse does not have severe muscle wasting in any muscle as seen in humans.

2.2 Is the natural history of the disease similar to human's regarding:

2.2.1 Time to onset;

Describe whether the model manifests the disease at around the same phase of development a human would. It is necessary to reference literature that compares the human and the species' development, which is available for the most prevalent animal model species, such as mice, rats and dogs.

> **How to score?**
>
> *Yes.*
> Although GRMD dogs have a somewhat high rate of neonatal death (~25%) not seen in either mdx mice or humans, the first clinical signs appear at 6-9 weeks, analogous to humans at 2-4 years.
>
> *No.*
> The first clinical signs appear at 15-20 weeks, much later than humans at 2-4 years.

2.2.2 Disease progression;

Describe whether the disease progresses in the model at around the same rate and leading to similar outcomes as in humans. The natural history of the disease shall be described comparatively in humans and the animal model, including the time to onset of different symptoms and complications.

How to score?

Yes, completely.
In humans, first symptoms start around 2.5 years, with loss of ambulation around 10 to 15 years. In GRMD dogs, muscular atrophy along with joint involvement (including jaw mobility), abnormal gait, decreased respiratory function and cardiomyopathy appears around six months of age. Like DMD, there is a honeymoon phase between 6 and ten months wherein the disease is relatively stable. Loss of ambulation is frequent, and like humans, death is commonly caused by heart failure and/or cardiomyopathy.

Yes, partially.
In humans, first symptoms start around 2.5 years, with loss of ambulation around 10 to 15 years. In GRMD dogs, muscular atrophy along with joint involvement (including jaw mobility), abnormal gait, decreased respiratory function and cardiomyopathy appears around six months of age. Like DMD, there is a honeymoon phase between 6 and ten months wherein the disease is relatively stable. Loss of ambulation is infrequent, with only around 1/3 of GRMD dogs losing ambulation completely, although they do lose mobility considerably.

No.
In humans, first symptoms start around 2.5 years, with loss of ambulation around 10 to 15 years. In GRMD dogs, muscular atrophy along with joint involvement (including jaw mobility), abnormal gait, decreased respiratory function and cardiomyopathy appears later in life, at around two years of age. There is no honeymoon phase wherein the disease is relatively stable. Loss of ambulation is infrequent, with only around 1 in 20 GRMD dogs losing ambulation completely, with most dogs being mobile until death.

2.2.3 Duration of symptoms;
Describe whether the symptoms are manifested for the same duration as the ones in humans (e.g. lifelong or temporary).

How to score?

Yes, completely.
A: Symptoms of DMD are lifelong due to permanent muscle degeneration. In mdx mice, once symptoms get worse, there is also permanent damage that leads to reduced lifespan.

Yes, partially.
A: Symptoms of DMD are lifelong due to permanent muscle degeneration. In mdx mice, once symptoms get worse until around 10 to 15 weeks, in which muscles sustain significant damage but do not further degenerate allowing some retention of function.

No.
A: Symptoms of DMD are lifelong due to permanent muscle degeneration. In mdx mice, symptoms get worse until around 10 to 15 weeks, in which muscles start to recover and partially regain their function.

2.2.4 Severity

Describe whether the severity of the symptoms manifested in the model is similar to the ones manifested in humans. This shall be done comparatively, with a brief description of both the human and animal situation.

How to score?

Yes, completely.
A: Mdx mice show the same level of muscle degeneration as seen in humans, including great variation in fibre size, architecture and continuous necrosis and proliferation of connective tissue.

Yes, partially.
A: Mdx mice do not show the same level of muscle degeneration as seen in humans, except for the diaphragm which shows progressive degeneration evident at six months of age with great variation in myofibre size, architecture and continuous necrosis and proliferation of connective tissue.

No.
A: Mdx mice do not show the same level of muscle degeneration as seen in humans in any muscle.

3. Genetic Validation

3.1 Does this species also have orthologous genes and/or proteins involved in the human disease?

Describe whether the primary genes and proteins involved in the pathophysiology of the disease in humans are present in the animal model species. Identification of such genes shall ideally be made by a working group of experts in the field. Alternatively, it can be done by searching for reviews on the genetic background of the disease and genome-wide association studies (GWAS). Once identified, the reasoning for such definition shall be justified with scientific literature. Additionally, for each identified gene, a brief description of its function and role in the pathophysiology of the disease along with its GeneID and location shall be reported for both humans and model species. The GeneID and location can be found at PubMed's database, available at https://www.ncbi.nlm.nih.gov/gene.

How to score?

For each relevant gene/protein, a subsection shall be created and individually analysed. The total score for this question shall be calculated by the sum of each subsection (which can be answered according to the pre-specified answers).

Examples
Yes.
3.1.1 Dystrophin: the human dystrophin gene (Gene ID: 1756) is the largest gene found in humans measuring a total of 2.4 Mb located at Xp21.2-p21.1. Dystrophin forms parts of the dystrophin-glycoprotein complex (DGC), responsible for connecting the inner cytoskeleton and the extracellular matrix. In dogs, the dystrophin gene (Gene ID: 606758) is located at chromosome X.

No.
3.1.1. The dystrophin gene together with its protein product is not present in GRMD dog muscles. The human dystrophin gene (Gene ID: 1756) is the largest gene found in humans measuring a total of 2.4 Mb located at Xp21.2-p21.1. Dystrophin forms parts of the dystrophin-glycoprotein complex (DGC), responsible for connecting the inner cytoskeleton and the extracellular matrix.

3.2 If so, are the relevant genetic mutations or alterations also present in the orthologous genes/proteins?

Describe the specific alterations (mutations, deletions etc.) in those genes and proteins present in humans and check whether they are also present in the model. Relevant genetic alterations (mutations, deletions, rearrangements etc.) shall be included for each gene identified in question 3.1. The procedure to identify such alterations follows the same principles, in which the identification of GWAS plays a more prominent role. A brief description of the effect of each alteration shall be included for each gene, contextualising it for both humans and the animal model.

How to score?

For each relevant gene/protein, a subsection shall be created and individually analysed. The total score for this question shall be calculated by the sum of each subsection (which can be answered according to the pre-specified answers).

Examples
Yes, completely.
A: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X. In that sense, DMD being an X-linked inherited disorder, the mdx mouse has the same aetiology (spontaneous mutations). This is caused by a nonsense point mutation (C-to-T transition) in exon 51 that aborts full-length dystrophin expression, which is the case 13% human patients and is the site of action of eteplirsen, an exon-skipping drug approved for the treatment of DMD.

Yes, partially.
A: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X. In that sense, DMD being an X-linked inherited disorder, the mdx mouse has the same aetiology (spontaneous mutations). Nevertheless, mdx mice have a nonsense point mutation (C-to-T transition) in exon 23 that aborts full-length dystrophin expression. This is not the case for human patients who have other alterations that lead to dystrophin deficiency.

No.
A: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X. The mdx mouse does not have the same mutation, as the lack of dystrophin in the muscles is caused by a deletion located at chromosome 2, completely different from the human situation.

3.3 If so, is the expression of such orthologous genes and/or proteins similar to the human condition?

Describe whether the expression (under-, normal or overexpression) levels of the relevant genes and proteins mentioned in the questions above and check whether the model has the same pattern of expression. If relevant, temporal (i.e. relevant only during a specific stage of development of disease phase) and spatial (i.e. present only in specific tissues) aspects of expression must also be considered.

How to score?
For each relevant gene/protein, a subsection shall be created and individually analysed. The total score for this question shall be calculated by the sum of each subsection (which can be answered according to the pre-specified answers).

Examples
Yes.
3.3.1 Dystrophin: Like DMD patients, mdx mice are dystrophin-deficient not expressing full-length dystrophin.

No.
3.3.1 Dystrophin: Mdx mice are not dystrophin-deficient, expressing fully functional dystrophin.

4. Biochemical Validation

4.1 If there are known pharmacodynamic (PD) biomarkers related to the pathophysiology of the disease, are they also present in the model?

Describe whether the known and well-established PD biomarkers relevant to the pathophysiology of the disease in humans are also present in the model. Pharmacodynamic biomarkers are defined as molecules which can describe the current state of a disease. Pharmacodynamic biomarkers can also be prognostic

biomarkers. Identification of such biomarkers shall ideally be done by a working group of experts in the field. Alternatively, reviews published in the literature can be used to identify these biomarkers. In both cases, scientific literature shall be included to support their choice. If relevant, temporal (i.e. relevant only during a specific stage of development of disease phase) and spatial (i.e. present only in specific tissues) aspects must also be considered.

How to score?
If more than one biomarker is identified, a subsection shall be created for each relevant biomarker and individually analysed. The total score for this question shall be calculated by the sum of each subsection (which can be answered according to the pre-specified answers).

Examples
Yes.
A: High levels of creatine kinase (CK) – a marker of muscle damage, with the muscle-type variants being predominant in plasma are also present in mdx mice.

No.
A: Creatine kinase (CK) levels – a marker of muscle damage, with the muscle-type variants being predominant in plasma are not elevated in mdx mice.

4.2 Do these PD biomarkers behave similarly to humans'?

Describe whether the behaviour of such biomarkers (higher or lower levels) is similar to the one seen in humans. If relevant, temporal (i.e. relevant only during a specific stage of development of disease phase) and spatial (i.e. present only in specific tissues) aspects of expression must also be considered.

How to score?
If more than one biomarker is identified, a subsection shall be created for each relevant biomarker and individually analysed. The total score for this question shall be calculated by the sum of each subsection (which can be answered according to the pre-specified answers).

Examples
Yes.
Like humans, ZDF rats have increased levels of blood glucose.

No.
Unlike humans, ZDF rats have normal levels of blood glucose.

4.3 If there are known prognostic biomarkers related to the pathophysiology of the disease, are they also present in the model?

Describe whether the known and well-established prognostic biomarkers in humans are also present in the model. Prognostic biomarkers are defined as any

measurement which can often predict the rate of progression of the disease and/or the likelihood of an outcome (e.g. recurrence). They can also be pharmacodynamic biomarkers. The identification of such biomarkers shall ideally be made by a working group of experts in the field. Alternatively, reviews published in the literature can be used to identify these biomarkers. In both cases, scientific literature shall be included to support their inclusion. If relevant, temporal (i.e. relevant only during a specific stage of development or disease phase) and spatial (i.e. present only in specific tissues) aspects must also be considered.

How to score?
If more than one biomarker is identified, a subsection shall be created for each relevant biomarker and individually analysed. The total score for this question shall be calculated by the sum of each subsection (which can be answered according to the pre-specified answers).

Examples
Yes.
A: Glycaemic markers are also prognostic markers as higher glycaemic levels can potentially lead to faster worsening of the diabetic condition. Glycaemic markers can also be measured in ZDF rats.

No.
A: Glycaemic markers are also prognostic markers as higher glycaemic levels can potentially lead to faster worsening of the diabetic condition. Unlike humans, no glycaemic markers can be measured in ZDF rats.

4.4 Do these prognostic biomarkers behave similarly to humans'?

Describe whether the behaviour of such biomarkers (higher or lower levels) is similar to the one seen in humans. If relevant, temporal (i.e. relevant only during a specific stage of development of disease phase) and spatial (i.e. present only in specific tissues) aspects of expression must also be considered.

How to score?
If more than one biomarker is identified, a subsection shall be created for each relevant biomarker and individually analysed. The total score for this question shall be calculated by the sum of each subsection (which can be answered according to the pre-specified answers).

Examples
Yes.
Like humans, ZDF rats have increased levels of blood glucose.

No.
Unlike humans, ZDF rats have normal levels of blood glucose.

5. Aetiological Validation

5.1 Is the aetiology of the disease similar to humans'?

Describe whether the aetiology of the disease in the model is similar to humans regarding both genetic and environmental (including lifestyle) factors. A brief review of what is known of the aetiology of the disease shall be included together with a comparative discussion on the animal model's disease aetiology. Genetic factors shall be cited in reference to the Genetic Validation domain, and environmental factors shall also be described.

It can be the case that a disease's aetiology is not yet known (idiopathic conditions) or there are many plausible causes (e.g. Alzheimer's Disease). In these cases, a discussion must be provided on the human aetiology including all current theories and a comparison of the model to each of them.

How to score?
If the disease's aetiology is not known or there are many concurrent hypotheses, since there is no 'correct' parameter to compare to, models which fit in one of the possible aetiologies shall be scored with a 'Yes, partially'.

Examples
Yes, completely.
A: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X. In that sense, DMD is an X-linked inherited disorder, the mdx mouse has the same aetiology (spontaneous mutations). This is caused by a nonsense point mutation (C-to-T transition) in exon 51 that aborts full-length dystrophin expression, which is the case 13% human patients and is the site of action of eteplirsen, an exon-skipping drug approved for the treatment of DMD.

Yes, partially.
A: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X. In that sense, DMD is an X-linked inherited disorder, the mdx mouse has the same aetiology (spontaneous mutations). Nevertheless, mdx mice have a nonsense point mutation (C-to-T transition) in exon 23 that aborts full-length dystrophin expression, which is not necessarily the case for human patients who can have other mutations that may also lead to dystrophin deficiency.

No.
A: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X. The mdx mouse does not have the same mutation, as the lack of dystrophin in the muscles is caused by a deletion located at chromosome 2, completely different from the human situation.

6. Histological Validation

6.1 Do the histopathological structures in relevant tissues resemble the ones found in humans?

Describe whether the main histopathological features in target tissues in humans are also present in the model. Identification of such structures shall ideally be made by a working group of experts in the field. Alternatively, reviews describing the disease and the animal model commonly make a comparison between histopathological features and can be used to identify such structures. A brief description of the comparison between the human and the animal situation shall be included.

How to score?
The total score for this question is calculated based on the proportionality of histopathological features modelled and partially modelled. For example, if there are three histopathological features defined for a given disease and a model simulates one histopathological feature completely (whole point), one partially (half a point) and the last one not at all, the score would be calculated by multiplying 1.5/3 by the weight of this question.

Examples
Histopathological feature modelled.
Muscle regeneration: existent with similar intensity as humans, being unable to outpace the increasing muscle necrosis.

Histopathological feature partially modelled.
Muscle regeneration: existent but at a somewhat faster pace than in humans, considerably slowing down the increasing muscle necrosis.

Histopathological feature not modelled.
Muscle regeneration: existent but at a much faster pace than in humans, effectively countering muscle necrosis with almost no effect on function.

---

7. Pharmacological Validation

7.1 Are effective drugs in humans also effective in this model?

Describe whether drugs proven to be effective in humans are also effective in the model, which indicates that the underlying mechanisms of the disease present in humans are also present in the model. The objective of this domain is to compare the relevance of pathways knowingly involved in the human disease's pathophysiology. Thus, different drug classes are used as examples to test the relevance of specific pathways (even if they overlap) in the model when compared to humans.

An effective drug is defined as a drug that has been determined to have a positive benefit-risk profile by the Committee for Medical Products for Human Use (CHMP), and thus received a Marketing Authorisation Application from the European Medicines Agency (EMA); or has been approved by the Food and Drug Administration (FDA). Drugs are grouped in classes by their mechanism of action and/or chemical similarity. Drug classes can be identified by searching for their Anatomical Therapeutic Chemical Classification System (ATC), publications of preclinical or clinical studies, reviews of management and/or therapy of the disease, consulting pharmacology textbooks and/or by accessing commercial drug databases such as Adis Insight and Pharmaprojects. All drugs approved for the treatment of a given indication shall be included. Inclusion of all drugs of each class allows a more systematic assessment of the pathway in the model for many reasons. The first is by preventing the introduction of bias by selectively including a drug which has positive or negative effects not shared by the rest of the class. Secondly, by skewing of results in case the model responds differently to a particular drug than to the rest of the class. Finally, by correcting for the fact that not all drugs of a class will be tested in all models.

The methodology (i.e. search string, database(s) searched, date of search and number of results found) to search for animal studies must be disclosed to allow reproducibility. Search strings for this question shall consist of the model's name (including abbreviations and plurals) and a drug (including drug codes and alternative names), e.g. *("ZDF rat" OR "ZDF rats" OR "Zucker Diabetic Fatty rat" OR "Zucker Diabetic Fatty rats") AND (exenatide OR "AC 2993" OR "AC 2993 LAR" OR Bydureon OR Byetta OR "Ex4 Peptide" OR "Exendin 4" OR Exendin-4 OR "ITCA 650").* In case too many hits are found, we advise to clean the results by adding [mesh], [tiab] to the model and drug search components.

Studies shall only be included if they include at least a control (placebo or an approved drug) and a monotherapy arm of the drug (combinations can be considered if they are the standard of care or if a combination drug is being developed). This does not apply to indications for which no approved treatment is available. Studies in which the treatment arm consists of plant extracts, phytotherapeutics and dietary supplements shall not be included unless they are extensively characterised. Studies shall be included only if the drug being tested is administered *in vivo* after the onset of the disease. Prevention studies (when a drug is administered before the onset of the disease) shall only be included if the indication aims to prevent the development of a condition. In case of Advanced Therapeutics Medicinal Products (ATMPs) or other drugs that are designed to interact specifically with human receptors or other structures, it is advised to also include in the search string a term for homologous drugs for that species. A summary of results shall also be included, highlighting the

FIMD: a standardised framework to validate, assess and compare models of disease

main effects of each drug primarily on the functional outcomes and secondarily on the other types of outcomes, according to the authors' significance testing. Conflicting results shall be reported and briefly discussed.

Additionally, a secondary file shall be created to include the classification of each study according to the concepts of general outcome and outcome context as defined below (How to score? Questions 7.1 and 7.2). A reporting quality and risk of bias assessment shall also be performed according to the parameters defined in the Reporting Quality and Risk of Bias Assessment in section F. An example is available online as S7 Supporting Information.

7.2 Are ineffective drugs in humans also ineffective in this model?

This question aims to identify whether drugs proven to be ineffective in humans are also ineffective in the model, indicating that the underlying mechanisms by which these drugs act have either similar relevance or are not present neither in humans nor the model.

Ineffective drugs are defined as drugs that due to lack of efficacy have either failed in clinical trials – phase II or III or were withdrawn by either agency; and do not share a mechanism of action with any drug currently approved by either EMA or FDA for the same indication. They can be identified by searching clinicaltrials.gov for clinical trials for any given indication. With the list of drugs (and drug codes) currently in development, drugs that have failed at phase II or III due to lack of efficacy can be identified via literature search of clinical trials published and/or press reports from originator/licensee companies. Additionally, a literature search can also be conducted for the mechanism of actions of drugs identified this way, so that even if a drug does not have clinical trials registered at clinicaltrials.gov, it could still be identified through its mechanism of action. When using the mechanism of action in the string, it is advised to look for the specific medical subject headings (MeSH) in Pubmed or Emtree terms in Embase as they allow the search for all drugs already catalogued within that class. Alternatively, the use of commercial drug databases such as Adis Insight and Pharmaprojects, which offer specific filters for terminated projects, is also encouraged. Drugs which have been withdrawn by regulatory agencies due to lack of efficacy can easily be identified by searching EMA and FDA's websites. Only drugs which have been tested in clinical trials that investigate efficacy shall be included.

This question follows the same reporting standards as question 7.1.

How to score? Questions 7.1 and 7.2.

The main findings of the included studies shall be reported in a brief description of each drug. The total score is calculated by the product of the weight of the pharmacological validation section by the sum of each subsection's score. The latter is calculated by the product of the general outcome (G_O), outcome context (O_c) and subsection value (S_v). All values shall be reported with four decimals. A formula for such calculation is provided below:

$$Total\ Score = W \sum_{Sj}^{Si} (G_0 \cdot O_c \cdot S_v)$$

The general outcome (G_O) refers to whether the studies included are in line with what was seen in the clinic: effective drugs are approved, and ineffective drugs have their development terminated or are withdrawn. This shall be only considered for functional outcomes (category II) as defined below. Other findings (e.g. histopathological changes) shall not be considered. Each study's outcome can be positive (Y), negative (N) or conflicting/partially in line (P). A study shall be classified as positive if it shows a statistically significant difference as reported by the authors in the functional outcome and negative if it did not. Studies which initially show a statistically significant difference and during the study lose such significance shall be classified as partially in line (P) unless such effect is characteristic of the drug/drug class being tested. Studies which report conflicting results in different functional outcomes analysed shall also be classified as (P). The score of each drug is calculated by the majority of studies, and it can be 0, 0.5 or 1. A score of 1 means that most studies are in line with the clinic, indicating the model could reliably predict human efficacy for that specific drug. A model which scores 1 in most drugs has most of the relevant pathways used for the treatment of a condition, meaning it is more likely to be able to predict human response to a drug involving those pathways.

Studies classified as category I (as defined below) are excluded as they, per definition, did not assess the functional outcomes used or comparable to the ones used in the clinical setting. In case a drug only has category I articles or has no published studies on the model being validated, the score shall be 10%, standing for 'unclear'. For example, if a drug has five studies in the model being validated, of which three were positive, one conflicting or partially positive and one negative, this multiplier's value is 1. If the number of studies with a positive and a negative assessment is the same or if most studies are conflicting, the section gets a multiplier with half of the points (0.5). If the majority is negative, the section gets a multiplier of 0.

The outcome context multiplier is calculated based on the weighted value of each study according to which category (I or II) a given study was classified. The classification in categories is based on the type of outcomes they report and is defined below:

Category I: Genetic/Biochemical/Histopathological (GBH): if an article only reports outcomes related to changes in gene expression, biochemical and/or histopathological marks of disease (e.g. inflammatory cell infiltration or blood levels of a commonly elevated enzyme), it shall be classified as category I. Articles in this category get a 0.5 multiplier for the outcome context score;

Category II: Functional: any article that reports at least one functional outcome shall be classified in category II. Functional outcomes are outcomes that can be considered 'clinically relevant' outcomes. These outcomes measure the function affected by the disease, offering the highest translational value for the clinical context. These outcomes are defined by the (mostly primary) outcomes used in pivotal trials (usually phases II/III) of approved drugs. If no drug is approved for a given indication, trials in the pipeline can be used as a reference. If there are no clinical trials for that indication, such outcomes can be defined in consultation with experts in the field. The justification for the selected outcomes must be provided (specifically in

question 8.1). Measures of muscle strength in degenerative muscle diseases (e.g. grip strength of fore- and hindlimbs in Duchenne Muscular Dystrophy) are examples of functional outcomes. Well-established surrogate outcomes that would otherwise be classified as genetic/biochemical/histopathological outcomes – such as HIV viral load for HIV infection/AIDS, glycaemic parameters (e.g. blood glucose, HbA1c, oral glucose tolerance test (OGTT) and urinary glucose) and cholesterol levels for diabetes and atherosclerosis, respectively – shall also be considered functional outcomes. Articles in this category get a 1.0 multiplier for the outcome context. Thus, if of 3 studies, 2 are category II and one category I, the outcome context multiplier is 2.5 (two full points plus one half-point) divided by 3 (total number of studies), which equals 0.8333.

The last multiplier is the subsection value, which is a simple division of 1 by the number of identified drug classes. An indication that has two drug classes approved (7.1.1 and 7.1.2, for instance), has a section value of 0.5. Within a subsection (i.e. each drug class), the subsection value is divided equally among all drugs. The final score (product of all multipliers) is then multiplied by the section's weight to achieve the final score. The same approach is taken for ineffective drugs.

7.3 Have drugs with different mechanisms of action and acting on different pathways been tested in this model? If so, which?

Describe whether the diversity of known mechanisms of action of drugs tested in humans was also tested in the model, providing a better characterisation of it.

How to score?
This question shall be answered by reporting the number of mechanisms of action of approved and failed drugs (in reference to questions 7.1 and 7.2) in comparison to the number of mechanisms of action assessed in the model being validated. The score of this question is calculated based on the proportionality of the mechanisms of action tested in the model. For instance, if there are five known mechanisms of action for which there are approved drugs and 3 for failed drugs, the total number of identified mechanisms of action is 8. If based on the answers in questions 7.1 and 7.2, 5 of these mechanisms have been tested on this model, the score is calculated by multiplying 5/8 by the weight of this question.

8. Endpoint Validation

8.1 Are the endpoints used in preclinical studies the same or translatable to the clinical endpoints?

Describe whether the endpoints used in at least one study included in the pharmacological validation are the same or can be translated to the clinical endpoints commonly used in clinical trials, indicating their feasibility in the model. Translatable endpoints can be defined as endpoints that measure the same or similar parameters but in a different context, often adapted to be performed in

animals, such as muscle strength (e.g. fore- or hindlimbs grip strength on rotarod in comparison with the execution of specific tasks that measure muscle strength in humans). Well-established surrogate endpoints (e.g. blood pressure in hypertension, viral load in HIV infection or blood glucose in diabetes) shall also be considered surrogate endpoints. A discussion on the translatability of such endpoints the justification for their classification in category I or II, as defined in question 7.1, must be provided. In addition to scientific literature, guidelines by regulatory agencies such as the FDA and EMA should be used to substantiate this discussion whenever relevant.

How to score?

Yes.

All but one study performed in ZDF with agents to treat type 2 diabetes have used glycaemic parameters (e.g. glycaemia, HbA1c, OGTT) or other measures of insulin sensitivity. These measurements also represent the primary outcomes of trials testing new drugs for the treatment of type 2 diabetes, which aim to control glycaemia.

No.

Most studies performed in ZDF with agents to treat type 2 diabetes focus on histopathological improvement of the pancreas and gene expression. Glycaemic parameters (e.g. glycaemia, HbA1c, OGTT) or other measures of insulin sensitivity, which are often the primary outcomes of trials testing new drugs for the treatment of type 2 diabetes, are not included.

8.2 Are the methods used to assess preclinical endpoints comparable to the ones used to assess related clinical endpoints?

Describe whether the methods used to assess these endpoints (e.g. biochemical techniques) are comparable to the ones used to assess such endpoints in the clinical setting. This shall also include an assessment of the state of mind (e.g. wakefulness and/or willingness) if relevant.

How to score?

Yes.

The measure of strength for both extension and flexion was taken in awake dogs, based on tasks they had to perform and were previously drilled. This is similar to DMD boys, who have to run a certain distance or perform certain tasks (e.g. get up from the floor without using hands).

No.

The measure of strength for both extension and flexion was taken in anesthetised dogs, in which flexion or extension was caused by electrical stimulation of muscles. DMD boys have to run a certain distance or perform certain tasks (e.g. get up the floor without using hands) while being awake and fully aware of the tasks.

D. Uncertainty Factor

The uncertainty factor is calculated as the percentage of the ratio of subsections answered with "Unclear" by the total number of subsections. This factor helps discriminate models which achieved a low final score because they are not well characterised. The number of subsections which had 'unclear' as an answer shall be divided by the total number of subsections.

If of 43 subsections, 15 were answered with 'unclear' either due to a lack of data or conflicting evidence, the uncertainty factor would be calculated as follows: 15/43*100 = 34.9%.

E. Similarity Factor

The similarity factor is calculated as the percentage of the ratio of subsections answered the same for two or more animal models being compared by the total number of subsections. This factor helps contextualise differences or the absence thereof in the scores of different animal models. The number of subsections which have the same answer in both models shall be divided by the total number of subsections. Answers can be 'Yes'; 'Yes, completely'; 'Yes, partially'; 'No'; or 'Unclear'.

For the SNH Validation question 2.1, it means having the same symptoms modelled, partially modelled or not modelled. The same applies to question 6.1 (Histological Validation). For the Pharmacological Validation questions 7.1 and 7.2, the same answer refers to the concordance with the clinical setting (General Outcome – G_o). For instance, if in model A, drug 1 had 20 articles in line with the clinical findings and one not, its G_o is 1.0. If in model B, drug 1 had 6 articles in line with the clinical findings and 3 not, its G_o is also 1.0. These subsections would then be considered as 'same answer' and would be counted for the calculation of the similarity factor. For question 7.3, the number of tested mechanisms of action must be the same.

An example of how to calculate the similarity factor based on the validated models of DMD is presented in Table E1. Calculation of the similarity factor for the mdx mouse and GRMD dog (see Chapter 3, Appendix E). Subsections from questions 2.1 and 6.1 are presented in the order they appear in the mdx mouse validation. Question/subsections with different answers are presented in bold.

TABLE E1 | Example of how to calculate the similarity factor per question. Parameters are presented in the same order they appear in the mdx mouse validation sheet (see Chapter 3, Appendix E).

Question (Subsections = 33)	Mdx mouse	GRMD dog
1.1	Yes, completely	Yes, completely
1.2	**Yes, partially**	**Yes, completely**
2.1.1 Reduce grown-up weight	Modelled	Modelled
2.1.2 ECG abnormality	Modelled	Modelled
2.1.3 Cardiomyopathy	Modelled	Modelled
2.1.4 Reduced lifespan	**Partially modelled**	**Modelled**
2.1.5 Muscle wasting	**Partially modelled**	**Modelled**
2.1.6 Cognitive and CNS defects	**Partially modelled**	**Unclear**
2.1.7 Loss of ambulation	**Not modelled**	**Partially modelled**
2.2.1	**No**	**Yes**
2.2.2	Yes, partially	Yes, partially
2.2.3	Yes, completely	Yes, completely
2.2.4	Yes, partially	Yes, partially
3.1.1 Dystrophin	Yes	Yes
3.1.2 Utrophin	Yes	Yes
3.2	Yes, partially	Yes, partially
3.3.1 Dystrophin	Yes, completely	Yes, completely
3.3.2 Utrophin	Yes, completely	Yes, completely
4.1	Yes, completely	Yes, completely
4.2	Yes, completely	Yes, completely
5.1	Yes, partially	Yes, partially
6.1.2 Limb muscle fibrosis	**Modelled**	**Partially modelled**
6.1.1 Muscle regeneration	**Not modelled**	**Modelled**
6.1.3 Adipose tissue	**Not modelled**	**Partially modelled**
7.1.1.1 Predinisone	$G_o = 1$	$G_o = 0.1$
7.1.1.2 Deflazacort	$G_o = 1$	$G_o = 0.1$
7.1.2 Ataluren	$G_o = 1$	$G_o = 0.1$
7.1.3 Eterplirsen	$G_o = 1$	$G_o = 0.1$
7.2.1 Bestatin	$G_o = 0.1$	$G_o = 0.1$
7.2.2 Stamulumab	$G_o = 0.1$	$G_o = 0.1$
7.3	**5**	**1**
8.1	**Yes**	**Yes**
8.2	**Yes, completely**	**No**
Total of # subsections	16	
Similarity Factor	16/33 = 48.5%	

F. Reporting Quality and Risk of Bias Assessment

A reporting quality and risk of bias assessment of all articles in domain *7. Pharmacological Validation* shall be included. The reporting quality and risk of bias assessment shall be based on an adaptation of the criteria defined by the ARRIVE guidelines and on SYRCLE's Risk of Bias Assessment Tool (1,2). A summary of the reporting quality and risk of bias assessments shall be provided per drug in the respective drug section (see Appendix E). When calculating the summary with all pharmacological studies, it is important to remove studies with more than one drug that has been administered to the same model to avoid duplication (see S7 Supporting Information).

While it is not our goal to focus on internal validity, we included the assessment of reporting quality and risk of bias to draw attention to the often poor internal validity of animal studies. It is up to the researchers to evaluate these assessments and their impact on the results of the studies of interest.

F1. Reporting Quality

It is important to note that the description of such parameters must be clear. For instance, 'standard' lab chow, housing or 'controlled' temperature and humidity are not considered as proper descriptions as well as 'carrying the experiments according to the regulations of a given institution'. Answers can be a 'Yes (Y)' if the paper properly described the parameter or 'No (N)' if it did not. The criteria for reporting quality are housing, husbandry, sample size, sample size calculation, acclimatisation, sex, background strain, blinding and randomisation.

Housing
1) type of facility: specify whether it is a regular facility or not (e.g. specific pathogen-free [SPF]);
2) type of cage or housing: describe the cage or housing (tank for fish) shape (if relevant) and material etc.;
3) bedding material: describe the material used in the bedding, if applicable;
4) number of cage companions: describe whether animals were individually housed or in groups and, if the latter, how big the groups were;

Husbandry
5) breeding programme: describe whether the animals were bred in-house or were bought from another company;
6) light/dark cycle: describe how many hours of light and dark the animals were exposed to;

7) temperature and humidity: describe both what the temperature and humidity were in the room were animals were housed;

8) quality of water etc (for fish): describe the quality parameters of the water;

9) type of food: describe what kind of food was provided, including the provider if bought commercially. If prepared in-house, describe the composition of the food;

10) access to food and water: describe whether animals could eat and drink *ad libitum* and if not, what were the periods animals could access food and water, and how much they could eat or drink;

11) environmental enrichment: describe whether there was any environmental enrichment (such as treadmills). If no environmental enrichment was provided, whether this was clearly reported;

Sample size

12) specify the total number of animals used in each experiment, and the number of animals in each experimental group;

Sample size calculation

13) explain how the number of animals was arrived at. Provide details of any sample size calculation used;

Any Blinding

14) describe whether blinding at any level and detail was described in the methodology;

Any Randomisation

15) describe whether randomisation at any level and detail was described in the methodology;

Acclimatisation

16) describe how long animals had to adapt to their facilities and/or training for functional measures;

Sex

17) describe whether animal sex was disclosed and if so, whether it included male (M), female (F) or male and female animals (B) in the treatment groups;

Background strain

18) whenever relevant, describe whether the background strain for the control was disclosed and which strain it was;

19) whenever relevant, describe whether the background strain for the model was disclosed and which strain it was;

F2. Risk of Bias (Signalling Question)

We suggest users to read and use of the Risk of Bias tool published by SYRCLE, which provides more information on its development and usability (2). Ideally, when assessing the risk of bias, users should take notes on the justification of their assessment. However, often the volume of studies will prevent the reporting of individual and summarised risks of bias at that level of detail. Thus, answers to each parameter follow the main signalling questions in which 'Yes (Y)' describes studies with a low risk of bias; 'Unclear (U)' studies in which there is not enough information to assess the risk of bias; and 'No (N)' studies with a high risk of bias. It is important to note that the additional signalling questions described in the original paper may help users conduct a proper risk of bias evaluation. The parameter referring to other risks of bias shall be reported with a brief explanation in the comments. The criteria for risk of bias are blinding, randomisation, sequence generation, baseline characteristics, incomplete outcome data, selective outcome reporting and 'other'. All criteria are defined below:

Blinding
20) allocation concealment (Was the allocation adequately concealed?): describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen before or during enrolment;
21) blinded outcome assessment (Was the outcome assessor blinded?): describe all measures used, if any, to blind outcome assessors from knowing which intervention each animal received. Provide any information relating to whether the intended blinding was effective;
22) blinded operations (Were the caregivers and/or investigators blinded from knowledge which intervention each animal received during the experiment?): Describe all measures used, if any, to blind trial caregivers and researchers from knowing which intervention each animal received. Provide any information relating to whether the intended blinding was effective.

Randomisation
23) random cage allocation (Were the animals randomly housed during the experiment?): if animals are not housed individually, whether the assignment of a cage to a treatment was randomised and how this was achieved;
24) random outcome assessment (Were animals selected at random for outcome assessment?): Describe whether animals were selected at random for outcome assessment, and which methods to select the animals, if any, were used;

Sequence Generation (Was the allocation sequence adequately generated and applied?)

    25) describe the methods used, if any, to generate the allocation sequence in sufficient detail to allow an assessment whether it should produce comparable groups;

Baseline Characteristics (Were the groups similar at baseline or were they adjusted for confounders in the analysis?)

    26) describe all the possible prognostic factors or animal characteristics, if any, that are compared in order to judge whether or not intervention and control groups were similar at the start of the experiment;

Incomplete Outcome Data (Were incomplete outcome data adequately addressed?)

    27) Describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomised animals), reasons for attrition or exclusions, and any re-inclusions in analyses for the review;

Selective Outcome Reporting (Are reports of the study free of selective outcome reporting?)

    28) State how selective outcome reporting was examined and what was found;

Other (Was the study apparently free of other problems that could result in high risk of bias?)

    29) State any important concerns about bias not covered by other domains in the tool.

References

1.   Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLOS Biology. 2010 Jun 29;8(6):e1000412.

2.   Hooijmans CR, Rovers MM, Vries RBM De, Leenaars M, Ritskes-hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC Medical Research Methodology. 2014;14(1):1–9.

FIMD: a standardised framework to validate, assess and compare models of disease

1) Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands;
2) Copernicus Institute of Sustainable Development, Innovation Studies, Utrecht University, Utrecht, The Netherlands;
3) Medicines Evaluation Board, Utrecht, The Netherlands.

3

Guilherme S. Ferreira[1],
Désirée H. Veening-Griffioen[1],
Wouter P. C. Boon[2],
Ellen H. M. Moors[2],
Christine C. Gispen-de Wied[3],
Huub Schellekens[1],
Peter J. K. van Meer[1, 3]
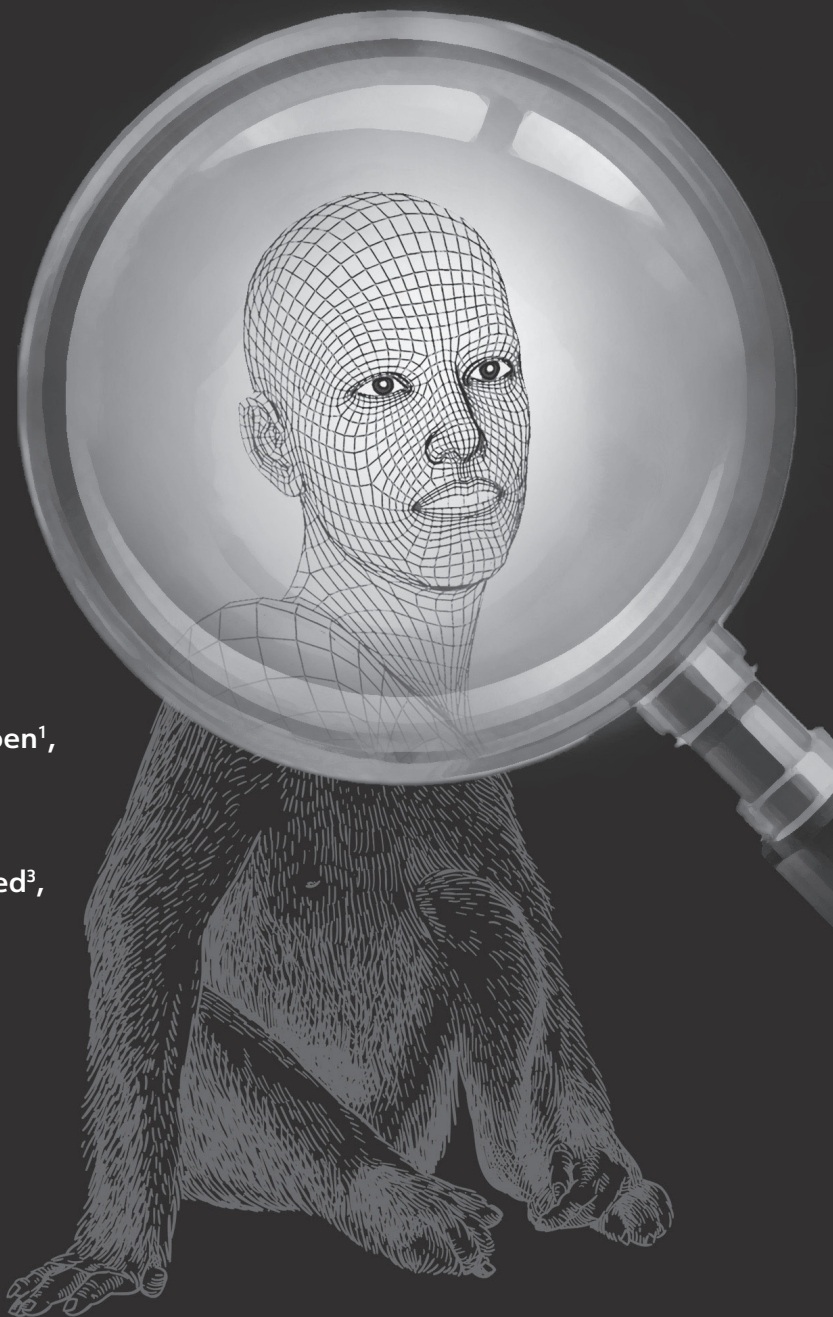
THE VALIDATION OF FIMD: INSIGHTS ON TYPE 2 DIABETES AND DUCHENNE MUSCULAR DYSTROPHY

Abstract

The Framework to Identify Models of Disease (FIMD) is a tool that allows researchers to assess, validate and compare animal models of disease. Here, we present FIMD's validation process, which consisted of two steps: a pilot study and complete validation. The pilot study was designed to test the feasibility of the framework's main features in a shorter timeframe. We selected two models of Type 2 Diabetes (T2D) for the pilot study owing to their extensive use in efficacy studies: the Zucker Diabetic Fatty (ZDF) rat and db/db mouse. The pilot study showed that there are only minor differences between the ZDF rat and the db/db mouse. However, a common point between these models was the lack of pharmacology studies. For the complete validation, we selected two models of Duchenne Muscular Dystrophy (DMD) based on their similarity to the human condition: the mdx mouse and Golden Retriever Muscular Dystrophy (GRMD) dog. The assessment with FIMD demonstrated there are significant differences between the mdx mouse and the GRMD dog in the Epidemiology, Symptomatology and Natural History (SNH) and Histology, Pharmacology and Endpoints domains. Our results indicate that the GRMD dog might be a more relevant model for drugs that target muscle degeneration. Moreover, we present a step-wise approach to validate animal models of disease using DMD as an example. Here, we show that FIMD can be used as a tool to systematically evaluate animal models of disease and therefore, allows the selection of models more likely to predict the human response.

Keywords: animal model, drug development, efficacy model, external validity, translational research, validation, 3Rs.

Introduction

The Framework to Identify Models of Disease (FIMD) consists of 21 questions across eight domains: Epidemiology, Symptomatology and Natural History (SNH), Genetics, Biochemistry, Aetiology, Histology, Pharmacology and Endpoints. The development of FIMD included an iterative validation of four animal models in two indications: Type 2 Diabetes (T2D) and Duchenne Muscular Dystrophy (DMD). We selected T2D because it has complex pathophysiology with a multifactorial aetiology and an extensive set of therapies is available (1). The Zucker Diabetic Fatty (ZDF) rat and the db/db mouse were chosen as T2D models owing to their routine use in drug screening for antidiabetic drugs. We chose DMD as an indication because it is caused by mutations in a single gene (known aetiology) and has few effective therapies available (2). The mdx mouse was chosen because it is the most commonly used animal model and the GRMD dog for better replicating the symptoms and histological features of DMD (3,4).

We adopted some methodological simplifications in the pilot study to quickly test the main features of the framework. Table 1 shows the differences between the pilot study and the complete validation. First, the search was conducted only on the PubMed database. Second, given the large number of studies in T2D, in question 7.1 (Pharmacological Validation), only one approved drug from each drug class was included. We favoured first-in-class drugs which are still in the market (either in the USA or EU) since it is likely that older drugs will have more studies published. If the first-in-class drug was withdrawn, the second-in-class was then preferred (and so on, if necessary). Including every first-in-class drug allowed an easy and clear standardisation as there can be only one first-in-class (unlike best- or last-in-class, which may be subjective and/or change over time). Finally, the reporting quality assessment consisted of a generic version of the parameters published in the Animal Research: Reporting In Vivo Experiments (ARRIVE) guidelines (5). For example, instead of explicitly assessing whether the animals were randomised into the treatment groups, the randomisation method was mentioned, or cage randomisation was performed, we assessed whether the authors provided any information on randomisation at all. The same was applied to blinding. For other multifaceted parameters (e.g. housing, husbandry), any facets mentioned warranted a positive evaluation.

The method for the complete validation was designed to minimise bias introduction. The complete validation includes searching on both PubMed and Embase, the largest biomedical research databases to ensure most animal studies are included. In the pharmacological validation section, all effective and ineffective drugs from all

drug classes must be included to avoid drug-specific effects being generalised to the whole drug class. We have made the reporting quality parameters more specific and unidimensional to ease their interpretation. Also, we added a risk of bias assessment with the SYstematic Review Center for Laboratory animal Experimentation (SYRCLE) tool (6) for all studies included in the Pharmacological Validation. We also provide a score calculator, which is available online[4]. These changes increased the robustness of the validation procedure and final product, the validation sheet.

In the following sections, we present the findings from the pilot study of T2D in the ZDF rat and db/db mouse; the complete validation of the mdx mouse and GRMD dog in DMD; and a step-wise approach to the use of FIMD as a tool to assess, validate and compare animal models of disease in drug development.

TABLE 1 | Differences in the methodology of the pilot study and the complete validation.

Parameter	Pilot Study	Complete Validation
Database	PubMed	PubMed and Embase
Pharmacological Validation Drugs	First-in-class	All drugs
Reporting Quality Assessment	8 generic parameters	19 specific parameters
Risk of Bias Assessment	None	10 parameters*

*From the SYRCLE's Risk of Bias tool (6).

Pilot study: insights on Type 2 Diabetes

A total of 195 publications was included for the ZDF rat and 282 for the db/db mouse. The final scores were 66.49 and 64.59, respectively. The relative scores per parameter are presented in Fig 1. There were only minor differences since the similarity factor between the two models is almost 90%. The ZDF rat has a few more studies characterising the expression of genetic alterations (e.g. KCNJ11) than the db/db mouse and therefore, scored slightly higher on the genetic validation section. Both T2D models have uncertainty factors above 20%, mostly because of the few studies included in the pharmacological validation. The similar aetiology in the rat and the mouse likely influenced the similar scores, but we explore these (dis)similarities further in **Chapter 4**. The validation sheets of the ZDF rat and db/db mouse are available in Appendix D.

---

[4] https://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0218014.s008

FIGURE 1 | T2D models results. Radar plot with the scores per parameter per model of T2D. The closer a parameter is to the edge, the better the model simulates that aspect of the human disease. SNH – Symptomatology and Natural History.

The reporting quality assessment was performed according to abridged parameters derived from the ARRIVE guidelines). Publications with experiments on two models were counted separately for each model while studies with more than one arm of an included drug were counted once. The reporting quality results (except for sex) for the ZDF rat and db/db mouse are presented in Table 2. Table 3 includes only the studies that reported the animal sex. The results per individual study for T2D[5] are available online. Most studies did not provide any information on the randomisation (66.3%, N = 251) or blinding (99.2%, N = 251). No publication explained how they arrived at the number of animals used in each experiment. Also, male animals were almost ten times more likely to be used than females while the use of both sexes is sporadic, making up only 4% of all studies. The reporting quality compliance was comparable in the ZDF rat and db/db mouse.

---

5 https://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0218014.s007

TABLE 2 | percentage of studies which comply totally or partially with reporting quality parameters adapted from the ARRIVE guidelines per model.

Parameter	ZDF rat (n = 96)	Db/db mouse (n = 169)	Total (n = 251)
	(%)		
Housing	36.3	34.2	35.1
Husbandry	84.4	78.3	80.9
Sample Size	91.2	93.2	92.8
Sample Size Calculation	0.0	0.0	0.0
Blinding	0.0	1.2	0.8
Randomisation	35.2	34.2	33.7
Acclimatisation	48.4	33.5	39.0
Sex	91.2	88.8	90.0
Background Strain	-	78.3	87.2

TABLE 3 | percentage of studies which included only male, female or both sexes per model.

Sex	ZDF rat (n = 83)	Db/db mouse (n = 143)	Total (n = 226)
	(%)		
Male	91.6	84.6	87.2
Female	2.4	12.6	8.8
Both	6.0	2.8	4.0

Complete validation: insights on Duchenne Muscular Dystrophy

A total of 58 articles were included for the mdx mouse and 41 for the GRMD dog. The final scores were 67.78 and 65.42 for the mdx mouse and GRMD dog, respectively. The relative scores per parameter are presented in Fig 3. The GRMD dog scores better in the epidemiological, SNH and histological domains while the mdx mouse does so in the pharmacological and endpoints domains. The GRMD dog mimics the natural history of the disease, symptoms (e.g. muscle wasting) and histopathological features (e.g. muscle regeneration) better than the mdx mouse. Especially for drugs which aim to slow down muscle degeneration or delay the disease onset, the GRMD dog is likely to generate more translatable data than the mdx mouse.

The difference in the pharmacological and endpoints domains stems mostly from the uncertainty factor, which is 18.2% and 6.1% for the GRMD dog and the mdx mouse, respectively. Since most drug screening studies are done in mdx mice, there are more studies available for the pharmacological validation. However, there are no published studies in GRMD dogs for most drugs tested in humans. The only published study that assessed a functional outcome did so in sedated dogs, which reduced the score of the Endpoints Validation. Hence, a comparison between these

models in these two domains is unlikely to be informative. The validation sheets of the mdx mouse and GRMD dog are available in Appendix E.
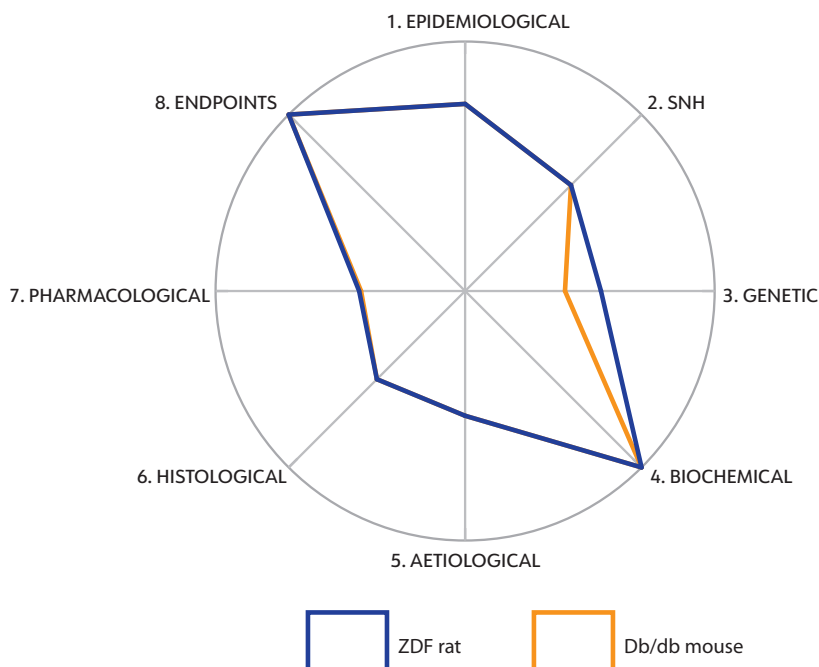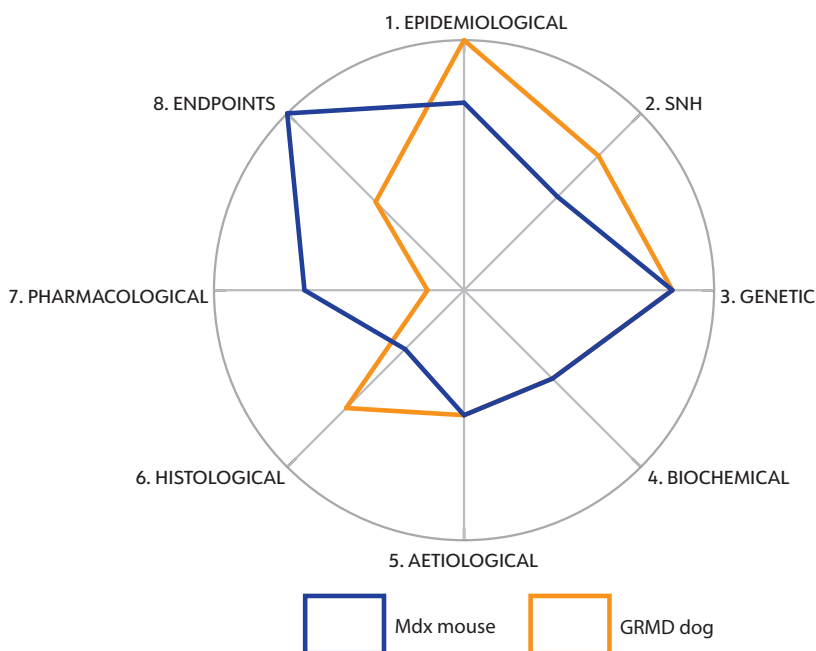

FIGURE 2 | DMD models results. Radar plot with the scores per parameter per model of DMD. The closer a parameter is to the edge, the better the model simulates that aspect of the human disease.

A reporting quality and risk of bias assessment was included based on parameters adapted from the ARRIVE guidelines (see Chapter 2, Appendix B, section F). Publications with experiments on two models were counted separately for each model. A total of 35 preclinical studies were included for the mdx mouse (32) and GRMD dog (3). Assessments of reporting quality and risk of bias per publication are provided online[6]. Table 4 shows the aggregated data for reporting quality, and Table 5 for the risk of bias. Most preclinical studies did not report important measures to reduce the risk of bias, such as any mention of randomisation (68.6%, N = 35) and blinding (57.1%, N = 35). Even studies which did mention these measures did not detail them enough to allow a proper evaluation of the risk of selection, performance and detection biases. The selective outcome reporting was the only parameter in which most studies were graded with a low risk of bias, indicating all outcomes defined in the methodology were reported in the results section.

---

[6] https://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0218014.s006

Male animals were more than two times more likely to be used than females, while studies including both sexes accounted for almost one-third of the total (31.8%, N = 35). Although in DMD this does not have a significant impact due to most patients with DMD being males, this trend is consistent with our findings for T2D, which show an overrepresentation of males in non-human animal studies.

TABLE 4 | Percentage of studies complying with the reporting quality parameters adapted from the ARRIVE guidelines per parameter per model.

Parameter	Mdx mouse (N =32)	GRMD dog (N = 3)	Total (N = 35)
	Y (%)		
Type of Facility	6.3	0.0	5.7
Type of Cage or Housing	12.5	0.0	11.4
Bedding Material	0.0	0.0	0.0
Number of Cage Companions	12.5	0.0	11.4
Breeding Programme	53.1	100.0	57.1
Light/Dark Cycle	25.0	0.0	22.9
Temperature and Humidity	3.1	0.0	2.9
Quality of the Water (fish)	-	-	-
Type of Food	6.3	0.0	5.7
Access to Food and Water	34.4	0.0	31.4
Environmental Enrichment	0.0	0.0	0.0
Any Blinding	43.8	33.3	42.9
Any Randomisation	34.4	0.0	31.4
Sample Size	84.4	100.0	85.7
Sample Size Calculation	6.3	33.3	8.6
Acclimatisation	18.8	0.0	17.1
Sex Disclosed	46.9	33.3	45.7
Male (N = 15)	53.3	0	50.0
Female (N = 3)	20.0	0	18.8
Both Male and Female (N = 16)	26.7	100	31.3
Background Control	-	-	-
Background Model	-	-	-

TABLE 5 | Percentage of studies with low (Yes) or unclear (Unclear) risk of bias per parameter per model.

Risk of Bias	Mdx mouse (N =32)	GRMD dog (N = 3)	Total (N = 35)
	Yes/Unclear (%)		
Allocation Concealment	0/100	0/100	0/100
Blinded Outcome Assessment	3.1/96.8	0/100	2.9/97.1
Blinded Operations	0/100	0/100	0/100
Random Cage Allocation	0/100	0/100	0/100
Random Outcome Assessment	0/100	0/100	0/100
Sequence Generation	0/96.8	0/100	0/97.1
Baseline Characteristics	3.1/96.8	0/100	2.9/97.1
Incomplete Outcome Data	18.8/67.7	0/100	17.1/71.4
Selective Outcome Reporting	96.9/3.2	100/0	97.1/2.9
Other	71.9/6.5	33.3/0.0	68.6/5.7

A step-wise guide to FIMD: the example of DMD

Based on the complete validation procedure, we elaborated a step-wise guide to executing FIMD using DMD as an example.

1) **Use MeSH and Emtree to identify all terminology related to the model and indication of interest, including plurals;**
   After an initial literature search to identify reviews models of DMD using the string '("animal model" OR "animal models") AND ("Duchenne Muscular Dystrophy" OR DMD) AND review', we selected the mdx mouse and the Golden Retriever Muscular Dystrophy (GRMD) dog. On PubMed, there are no MeSH terms for either the mdx mouse or the GRMD dog, so we opted for the use of 'mdx' and 'GRMD' as general, broad terms of the subsequent searches. In Embase, we used the terms from the Emtree (see Chapter 2, Appendix B, questions 7.1 and 7.2).

2) **Organise an expert group to define the disease parameters. If that is not possible, search for reviews on the human disease and animal models for that condition. Whenever necessary, justify the choice of disease parameters;**
   We could not set up an expert group for the definitions of the disease parameters. We started by using the terms identified in the previous item to search for general reviews on the models. We found most of the effective and ineffective drugs in these reviews combined with searches in ClinitralTrials.gov and the free version of AdisInsight. Using reviews' reference lists and forward citations, we identified most papers used in the validation. Since we could not

find any prognostic biomarkers for DMD, we decided to exclude questions 4.3 and 4.4 (both related to prognostic biomarkers), adding a justification and the references on which we based our decision.

3) **Answer the questions in the validation sheet template (Chapter 2, Appendix C) based on the information found during the search for papers. If, in the end, some questions are still unanswered, use specific search strings for these questions;**
Often papers will have data that can be used to answer more than one question. Just by reading reviews, their references and using forward citation on Google Scholar, we could answer all questions from FIMD. We did not have to use any specific search strings.

4) **For the Pharmacological Validation, assess each paper based on the instructions for questions 7.1, 7.2; and Chapter 2, Appendix B, sections F1. Reporting Quality and F2. Risk of Bias. Fill this information in a separate spreadsheet file (such as the excel sheet provided online[7]);**
We assessed 32 articles for the Pharmacological Validation of the mdx mouse and 3 for the GRMD dog. A report per paper, as well as aggregated data, is provided in the online version. A summary of the reporting quality and risk of bias assessment was provided for each drug (see Tables 3 and 4).

5) **Fill in the Score Calculator (available online) to get the total score and the radar plot;**
All the answers were filled in the score calculator. We had to adjust the calculation of the final score to account for the exclusion of questions 4.3 and 4.4, resulting in a total max score of 93.75. We calculated the equivalent score to 100 (), which for the mdx mouse was 67.78 and for GRMD dog was 65.41. The radar plot is automatically generated once all fields are filled out.

6) **Calculate the Uncertainty Factor according to Chapter 2, Appendix B, section D;**
More details on how to calculate the uncertainty factor are provided in Chapter 2, Appendix B, section D. The mdx mouse had an uncertainty factor of 6.1% while the GRMD dog had 18.2%. In both models, all 'unclear' answers were given to sections in the Pharmacological Validation, especially for the GRMD dog, which had only three studies with one of the five drugs identified.

---

[7] https://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0218014.s006

7) **If another model has been validated for the same indication, calculate the Similarity Factor according to Chapter 2, Appendix B, section E;**
   More details on how to calculate the similarity factor are provided in Appendix B, section E. The similarity factor was 48.5%, meaning that less than half of the questions were answered in the same way, which indicates there are considerable differences between the mdx mouse and the GRMD dog.

8) **Finalising the validation sheet;**
   Once the final score, the uncertainty and similarity factors are calculated, fill in this information in the first page of the validation sheet. Include the validation date (latest date of publication of the validation references). Determine the validation level as defined in Table 3. Describe the historical background of the model, as explained in Chapter 2, Appendix B. Fill each question's score.

Final considerations

The pilot study showed only minor differences between the ZDF rat and db/db mouse. Both animal models seem comparable across all eight domains. On the other hand, the findings of the complete validation of DMD models suggest the mdx mouse and GRMD dog have significant dissimilarities. The GRMD dog scores better than the mdx mouse in critical domains, such as SNH and Histology, but lacks enough studies for an adequate comparison in the Pharmacology domain. The scores in both the pilot study and the full validation suggest more sensitivity is needed, possibly through the adoption of an indication-specific weighting and scoring system. As discussed in the previous chapter, the development of such a system will require widespread adoption and gathering of data to calculate the weights based on the disease characteristics. Moreover, our findings regarding reporting quality and risk of bias were consistent across indications, and are in line with the literature (7–11). The poor reporting – and arguably performance – of the studies included in the pharmacological validation put their reliability into question.

Here, we showcased the feasibility of assessing and comparing animal models of disease with FIMD. The step-wise approach also provides researchers with a quick guide on how to start a full validation and use all features of the framework. Thus, researchers now have a tool to collect, aggregate and analyse disease model data in a systematic, reproducible and evidence-based fashion, allowing the selection of more predictive animal models.

References

1.	Gourgari E, Wilhelm EE, Hassanzadeh H, Aroda VR, Shoulson I. A comprehensive review of the FDA-approved labels of diabetes drugs: Indications, safety, and emerging cardiovascular safety data. Journal of Diabetes and its Complications. 2017;31(12):1719–27.
2.	Reinig AM, Mirzaei S, Berlau DJ. Advances in the Treatment of Duchenne Muscular Dystrophy: New and Emerging Pharmacotherapies. Pharmacotherapy. 2017;37(4):492–9.
3.	McGreevy JW, Hakim CH, McIntosh MA, Duan D. Animal models of Duchenne muscular dystrophy: from basic mechanisms to gene therapy. Dis Model Mech. 2015;8(3):195–213.
4.	Yu X, Bao B, Echigoya Y, Yokota T. Dystrophin-deficient large animal models: Translational research and exon skipping. American Journal of Translational Research. 2015;7(8):1314–31.
5.	Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLoS Biol. 2010 Jun 29;8(6):e1000412.
6.	Hooijmans CR, Rovers MM, de Vries RBM, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol. 2014 Mar 26;14:43.
7.	Bebarta V, Luyten D, Heard K. Emergency medicine animal research: Does use of randomization and blinding affect the results? Academic Emergency Medicine. 2003;10(6):684–7.
8.	Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans? BMJ : British Medical Journal. 2004;328(7438):514–7.
9.	Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ : British Medical Journal. 2007;334(7586):197–197.
10.	van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can Animal Models of Disease Reliably Inform Human Studies? PLoS Medicine. 2010 Mar 30;7(3):1–8.
11.	Peers IS, South MC, Ceuppens PR, Bright JD, Pilling E. Can you trust your animal study data? Nature Reviews Drug Discovery. 2014;13(7):560.

Appendix D - Pilot Study – Type 2 Diabetes (T2D)

MODEL NAME	Zucker Diabetic Fatty (ZDF)-*Lepr^fa*/Crl rat
INDICATION	Type 2 Diabetes Mellitus (T2D)
VALIDATION DATE	18.4.2017
TOTAL SUBSECTIONS	59
TOTAL SCORE	68.36
UNCERTAINTY FACTOR (%)	20.3

HISTORICAL BACKGROUND
The Zucker Fatty (ZF) rat was identified after crossing Merck M-strain with Sherman rats in 1961 (1,2). In these mice, the leptin receptor is mutated (the *fatty* or *fa* mutation). Another mutation – an autosomal recessive defect in pancreatic β-cells transcription inherited independently from the leptin mutation – occurred in a colony of outbred ZF rats, leading to the characterisation of a substrain with a diabetogenic phenotype called Zucker Diabetic Fatty (ZDF) rat (3). Further establishment of the line has been done by inbreeding rats with diabetic lineage (3).

1. EPIDEMIOLOGICAL VALIDATION	
1.1 Is the model able to simulate the disease in the relevant sexes?	Score
Yes, completely. **Remarks**: There are reports of difficulties in being able to properly simulate diabetes in female ZDF rats, which are commonly used as controls to males in preclinical studies (4,5). However, both Charles River Laboratory (the main supplier of this strain) and Corsetti et al have successfully induced the diabetic state in females by using specific diets (DI2468 and 48% high-fat diet, respectively) (3,5,6). Furthermore, this specific female resistance to the development of diabetes is in line with what is seen in humans, due to a possible protective effect of oestrogen on pancreatic beta-cells (7,8).	2
1.2 Is the model able to simulate the disease in the relevant age groups (juvenile, adult or ageing)?	Score
Yes, partially. **Remarks**: Although the time to onset of diabetes has been decreasing in the past years, ZDF rat still develops diabetes somewhat earlier than humans (9). However, it progresses into adulthood and ageing phases, similarly to the human disease (6,10–14).	1

2. SYMPTOMATOLOGY AND NATURAL HISTORY (SNH) VALIDATION	
2.1 Is the model able to replicate the symptoms commonly present in this disease? If so, which ones?	Score
Yes, completely. The most common symptoms of diabetes type 2 are polyphagia, polydipsia, polyuria, weight loss, fatigue, healing impairment and obesity (15). The latter may not be considered a symptom per se, being better classified as a co-morbidity. However, due to the high prevalence of this condition in T2D patients and likely association to the pathophysiology, it was also included in this section. Most of these symptoms are directly connected to the hyperglycaemia and dyslipidaemia and therefore they were translated into the following parameters: hyperglycaemia, hyperinsulinemia followed by hypoinsulinemia, dyslipidaemia (divided into hypercholesterolemia and hypertrigceridemia according to the common human pathological profile) and obesity.	13

2.1.1 Symptoms modelled

Hyperglycaemia: At 5-7 weeks, fasting blood glucose is already higher than lean littermates and with a further increase at 10-12 weeks and stabilising at a few-fold more than lean rats (11,14,16–23).

Hyperinsulinemia followed by hypoinsulinemia: insulin levels in ZDF rats are initially significantly higher than their lean littermates, however, they decrease steadily with age due to β-cell insufficiency (6,11,17,21).

Hypercholesterolemia: high cholesterol levels are found in ZDF rats as well as humans, which eventually leads to atherosclerosis (17,21,22).

Hypertriglyceridemia: triglyceride and free fatty acids (FFA) levels in ZDF rats are significantly higher than in lean littermates (11,14,17,19,21,22).

Obesity: ZDF rats are significantly heavier than their lean littermates at a young age, but tend to significantly lose weight after overt diabetes, which is consistent with the sudden weight loss seen in humans (11,16,18–21,23,24).

2.2 Is the natural history of the disease similar to humans regarding:	
2.2.1 Time to onset;	Score

	Score
No. **Remarks**: type 2 diabetes is usually diagnosed around 40 years, although age at diagnosis has been decreasing (9). The onset of diabetes in male ZDF rats may vary from 6 to 12 weeks, but commonly happening between 7 and 8 weeks (6,10–14).This happens at quite an early stage of life when compared to humans: rats 24 weeks of age are in a similar stage of development as humans around 18 years (25).	0

2.2.2 Disease progression;	Score

	Score
Yes, partially. **Remarks**: Diabetes progression in ZDF rats follows much of the human disease. The increase in sugar/fat intake in diets leads to an increase in insulin secretion, reducing the ability to activate the TK receptor (between 3 and 8 weeks of age). Consequently, this then leads to an increase in glycaemia and β-cell degeneration, finally leading to β-cell apoptosis and glucose intolerance (between 6 and 12 weeks of age) (6,10–13). The whole process, however, happens at quite an early stage of life when compared to humans: rats 6 months of age are in a similar stage of development as humans around 18 years (25). Although as mentioned before, the age of onset of diabetes has been decreasing in the past years (9). As for common diabetic complications, such as nephropathy, retinopathy and neuropathy, ZDF rats can partially reproduce them. They are able to develop renal lesions (such as albuminuria, glomerulosclerosis, tubulointerstitial scarring and inflammation) progressively with age, but these are confounded by natural ageing lesions and hydronephrosis (26–28). ZDF rats can also develop ocular lesions typical of diabetes such as cataract formation and apoptotic death of lens cells but lack other lesions commonly present in humans, such as pericyte degeneration, microaneurysms, and acellular capillaries (29–31). Likewise, ZDF rats can partially develop neurological complications, such as decreased caudal motor nerve conduction velocity (MNCV), impairment of acetylcholine-mediated vascular extension of epineural arterioles of the sciatic nerve, but do not show sympathetic neuronal dystrophy (24,32,33).	1

2.2.3 Duration of symptoms;	Score

	Score
Yes, completely. **Remarks**: like humans, once insulin resistance and glucose intolerance are set, β-cell apoptosis follows, worsening the diabetic condition progressively and leading to overt diabetes and diabetic complications (16,34).	2

2.2.4 Severity.	Score
Yes, partially. **Remarks**: the biochemical parameters in ZDF rats are similar to humans in severity, being considerably increased (please see section 2.1). However, the complications that follow after the onset of overt diabetes are somewhat milder in ZDF rats (lack of pericyte degeneration in the retinopathy or sympathetic neuronal dystrophy in nephropathy).	1

3. GENETIC VALIDATION	
3.1 Does this species also have orthologous genes and/or proteins involved in the human disease? If so, which?	Score
Yes, completely.

Although is widely accepted that diabetes type 2 has genetic factors involved, reliable tracking and identification of such genes remain a challenge (35–42). This is often referred to as the 'missing heritability' (36). In some cases, T2D can be caused by a single gene (such as mutations on HNF4A for Maturity-Onset-Diabetes of the Young type 1 – MODY1 – or PTF1A for neonatal diabetes) (35,37). Nonetheless, in most cases it is a polygenic and heterogeneous disease, meaning multiple genes might be involved and different combinations of polymorphisms can lead to the pathological phenotype, which makes the genetic validation of animal models at best extremely challenging (35,37,39–42).

With the advent of Genome-Wide Association Studies (GWAS), many genes have been associated with an increased risk of developing type 2 diabetes, shedding light on the complex genetic architecture of the disease (35,43,44). However, three of them have been consistently established in GWAS as likely genetic factors: TCF7L2, KCNJ11 and PPARG (35,38,45). Therefore, these were the only ones included in this section.
All three genes are present in ZDF rats (46–48). | 8 |

3.1.1 TCF7L2
TCF7L2 is a gene (also known as also known as TCF-4 or β-catenin interacting protein) located at 10q25.2-q25.3 in humans and at 1q55 in rats that codes a transcription factor involved in the WNT signalling pathway (46,49–52). This is the strongest and most well-replicated genetic factor associated to T2D (35,38,43,44,53,54). It acts as a nuclear receptor for β-catenin, which is involved in the secretion of GLP-1 in gut endocrine cells and various other genes (50,51,55). There are also reports that indicate a possible role in the incretin axis, adipocyte function and glucose production by the liver, despite some of the results being conflicting (52,56–60).

3.1.2 KCNJ11
KCNJ11 is a gene (also known as KIR6.2) located at 11p15.1 in humans and at 1q22 in rats that has been associated with an increased risk of developing T2D (43–45,48,61). It codes for a major subunit of the ATP-sensitive K^+ channel, an inward-rectifier potassium ion channel present in the pancreatic islets with direct influence over insulin secretion (62,63). Mutations in this gene are connected to neonatal diabetes (35).

3.1.3 PPARG
PPARG is a gene that codes for the transcription factor peroxisome proliferator-activated receptor-γ (PPARγ) located at 3p25.2 and at 4q42 in rats in humans (47,64). The PPARG is involved in adipocyte regulation, fat accumulation and glucose metabolism and it is stimulated by insulin in a ligand-dependent manner (65–68). Two isoforms exist in humans: PPARγ1 and PPARγ2, the first being expressed in most tissues and the second one mostly in the liver and adipose tissue (69). The isoform 2 is also the target of thiazolidinediones, such as rosiglitazone.

3.2 If so, are the relevant genetic mutations or alterations also present in the orthologous genes/proteins?	
Unclear.	0.5

3.2.1 TCF7L2

Unclear.

In humans, the rs12255372 and rs7903146 alleles are the most strongly associated variants with carriers of such type having reduced insulin secretion but not increased insulin resistance (70). Being homozygous for the high-risk allele doubles the chance of developing T2D (35). No studies investigating specific genetic alterations in this gene could be found in PubMed in ZDF rats using the string ("ZDF rat" OR "ZDF rats" OR "Zucker Diabetic Fatty rat" OR "Zucker Diabetic Fatty rats") AND TCF7L2 on 13/04/2017.

3.2.2 KCNJ11

Unclear.

In humans, the commonest polymorphism is a glutamate to lysine substitution at position 23 (E23K or rs5219) (71–74). This mutation is linked to a reduction in channel sensitivity, increasing the signal threshold for the release of insulin and impairing serum insulin response (62,71,72,75). Homozygous carriers of this mutation (KK) have an almost double risk of developing T2D when compared to non-carriers (73).

No studies investigating specific genetic alterations in this gene could be found in PubMed in ZDF rats using the string ("ZDF rat" OR "ZDF rats" OR "Zucker Diabetic Fatty rat" OR "Zucker Diabetic Fatty rats") AND (KCNJ11 OR KIR6.2 OR "KIR 6.2") on 13/04/2017.

3.2.3 PPARG

Unclear.

In humans, the SNP rs1801282 consists of a proline substitution for alanine in position 12 (Pro12Ala or P12A) of PPARγ2 (the other isoform is not affected), the proline allele is associated with an increased risk of developing T2D, higher BMI and decreased insulin sensitivity while the alanine one confers resistance (43–45,73,74,76–78). Even though the mechanism is still unclear, the Ala variant has a lower transactivation efficiency, reducing stimulation of PPARG target genes and therefore also reducing the levels of adipose tissue mass accumulation (76). Nonetheless, there are also reports of increased risk of cardiovascular disease in carriers of this polymorphism (79,80).

No studies investigating specific genetic alterations in this gene could be found in PubMed in ZDF rats using the string (PPARgamma) AND ("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") on 18/04/2017.

3.3 If so, is the expression of such orthologous genes and/or proteins similar to the human condition?	Score
Yes, partially.	1.07

3.3.1 TCF7L2

Unclear.

Two studies found the TCF7L2 gene is overexpressed in carriers of the risk genotype in pancreatic islets of T2D patients, but no changes in TCF7L2 expression were found by study of the Diabetes Genome Anatomy Project (DGAP) (35,56,81). One study by Shu and colleagues showed a possible protective effect on the islets against glucose- and cytokine-induced apoptosis and function impairment. This study was later retracted due to image duplication and concerns over data reliability (82). One study in ZDF rats has shown overexpression of TCF7L2 in the islets, unaffected by the blockade of SREBP-1c (83).

3.3.2 KCNJ11
Yes, completely.
KCNJ11 is underexpressed in human islets and ZDF rats islets and hypothalamus, which is consistent with impaired β-cell insulin release (84–87).
3.3.3 PPARG
Yes, partially. PPARG is overexpressed in adipose tissue in diabetic humans and ZDF rats, but expression levels in the liver are conflicting: underexpression being reported in humans and overexpression in ZDF rats (35,87).

4. BIOCHEMICAL VALIDATION	
4.1 If there are known pharmacodynamic (PD) biomarkers related to the pathophysiology of the disease, are they also present in the model?	Score
Yes, completely. **Remarks**: Glycaemic markers (blood glucose and HbA1c), cholesterol and triglycerides levels are directly related to the development of diabetes and diabetic complications, being used as PD biomarkers.	3
4.1.1 Glycaemic markers (blood glucose, HbA1c)	
All glycaemic markers can be measured in ZDF rats.	
4.1.2 Cholesterol	
Cholesterol levels can be measured in ZDF rats.	
4.1.3 Triglycerides	
Triglycerides levels can be measured in ZDF rats.	
4.2 Do these PD biomarkers behave similarly to humans'?	Score
Yes, completely.	2
4.2.1 Glycaemic markers (blood glucose, HbA1c)	
All glycaemic markers are increased in ZDF rats (see section 2.1.1).	
4.2.2 Cholesterol	
Cholesterol levels are increased in ZDF rats (see section 2.1.1).	
4.2.3 Triglycerides	
Triglycerides levels are increased in ZDF rats (see question 2.1.1).	
4.3 If there are known prognostic biomarkers related to the pathophysiology of the disease, are they also present in the model?	Score
Yes, completely. **Remarks**: Glycaemic markers are also prognostic markers as higher glycaemic levels can potentially lead to faster worsening of the diabetic condition (88–90). The same for cholesterol and triglycerides, which are directly involved in cardiovascular damage and diabetic complications (91,92).	3
4.3.1 Glycaemic markers (blood glucose, HbA1c)	
All glycaemic markers can be measured in ZDF rats.	
4.3.2 Cholesterol	
Cholesterol levels can be measured in ZDF rats.	

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

4.3.3 Triglycerides	
Triglycerides levels can be measured in ZDF rats.	
4.4 Do these prognostic biomarkers behave similarly to humans'?	Score
Yes, completely. **Remarks**: Like humans, ZDF rats have increased levels of blood glucose, cholesterol and triglycerides.	2
4.4.1 Glycaemic markers (blood glucose, HbA1c)	
All glycaemic markers are increased in ZDF rats (see section 2.1.1).	
4.4.2 Cholesterol	
Cholesterol levels are increased in ZDF rats (see section 2.1.1).	
4.4.3 Triglycerides	
Triglycerides levels are increased in ZDF rats (see question 2.1.1).	

5. AETIOLOGICAL VALIDATION	
5.1 Is the aetiology of the disease similar to humans'?	Score
Yes, partially. **Remarks**: diabetes type 2 is strongly associated with obesity, high-fat/carbs diet and sedentary lifestyle (9,93). Lifestyle changes such as a healthful diet, BMI control and reduction of smoking/alcohol intake could prevent as much as 90% of type diabetes cases (93,94). In ZDF rats, diabetes is caused by the concomitant presence of two different genetic factors. The first is a mutation in the extracellular domain of the leptin receptor (*lepr*) gene. Leptin is a hormone that can be produced by various tissues, but mostly by mature adipocytes in white adipose tissue. It has an important role in satiety when it is taken up by the brain from the circulation (34). The mutation in the leptin gene is called 'fa' because it was originally identified in the Zucker Fatty rats, from which the ZDF rat strain was derived (34). ZDF rats are homozygous for this missense mutation in the leptin receptor from A to C at position 880, which causes a Gln to Pro change in all identified isoforms of the Ob-R protein, while lean littermates do not have it (34,95). The second factor is an autosomal recessive defect in β-cell transcription, which is inherited independently from the *lepr* mutation and contributes directly to the manifestation of overt diabetes (96). In humans, although leptin has a similar function, it seems to be less critical for the regulation of energy expenditure as inactivating mutations in its receptor have milder effects than in rodents (97). Several genome-wide association studies (GWAS) have conflicting results regarding the association of polymorphisms in the *lepr* gene and risk of diabetes (98–100). Nevertheless, the existence of genetic factors is widely accepted, although they probably result from the product of several small to moderate gene effects (35–42). For overt diabetes to develop in humans, it is necessary to have a genetic predisposition, β-cell dysfunction and insulin resistance (usually acquired due to obesity). In that sense, ZDF rats are similar (96). Male ZDF rats develop diabetes even on low-fat diets while females need high-fat diets to do so, the latter being closer to the common human aetiology (6). This fact is also in line with the specific female resistance to the development of diabetes seen in humans, due to a possible protective effect of oestrogen on pancreatic beta-cells (7,8).	7

6. HISTOLOGICAL VALIDATION	
6.1 Do the histopathological structures in relevant tissues resemble the ones found in humans?	Score
Yes, partially. There is some evidence of β-cell mass reduction and pathogenic role of amyloid deposits in human patients, even though conflicting results to these notions are also present (101–104). Nonetheless, they were included to better characterise the models in this sheet. The pancreas was the only tissue selected for the histological validation because it primarily affects the pancreatic islets, in which long-term damage then leads to diabetic complications.	5.5
6.1.1 Histopathological features modelled	
B-cell mass: as seen in diabetic patients, the β-cell mass in ZDF rats islets is severely diminished with the development of overt diabetes (16,105).	
6.1.2 Histopathological features partially modelled	
Islet morphology: islets become irregularly delineated with the advancement of diabetes, also increasing in size and presenting vacuolations between the mantles. Even though the basic morphology is different than the human pancreas, the presence of fibrosis and fat infiltration is similar. Vacuolations commonly present in ZDF rat pancreas tissue are present in humans only due to tissue processing errors (12,16,102,106–108).	
6.1.3 Histopathological features not modelled	
Amyloidosis: although there is evidence of hypersecretion of islet amyloid polypeptide (IAPP) or amylin in ZDF rats' pancreas, there is no report of formation of the characteristic amyloid plaques as seen in humans (20,108,109).	

7. PHARMACOLOGICAL VALIDATION	
7.1 Are effective drugs in humans also effective in this model?	Score
Yes, partially. **Remarks**: one drug from each class approved by FDA and/or EMA for the treatment of type 2 diabetes was included.	4.02
7.1.1 Amylin Agonist: Pramlintide	
Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND pramlintide' on 27/02/2017, yielding no results.	
7.1.2 Biguanide: Metformin (110–131)	
Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND metformin' on 27/02/2017, yielding 26 results which were screened by title and abstract. A study retrieved through the search string for rosiglitazone with a metformin monotherapy arm was also screened. Of these, twenty-two (22) articles were included in this section.	
Results: most studies show a reduction of glycaemia measures, such as plasma glucose, HbA1c, OGTT; and triglycerides and free fatty acids. Studies which did not report a reduction in glycaemic parameters often cite dose as the probable reason for this finding. Metformin improves islet histology and increases β-cell total mass (increases expression of anti-apoptotic genes while decreasing pro-apoptotic ones) when compared to controls. It has also been shown to have a moderate effect on liver steatosis and to improve nephropathy by reducing genomic damage. However, it caused no improvement in the endothelial function. All but one article were classified as category II. Of the 21 studies classified as category II, fourteen (14) are in line with clinical findings.	

Quality assessment: according to the pre-specified criteria and n = 22, 41% have information on housing; 95% on husbandry, 91% on sample size, none on sample size calculation nor blinding, 50% on randomisation; 59% on acclimatisation; and 95% on sex (19 used only male animals, one used only female animals and one used both).

7.1.3 Bile Acid Sequestrant: Colesevelam (132)

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND colesevelam' on 27/02/2017, yielding 1 result which was screened by title and abstract. This article was included in this section.

Results: the only study shows a positive effect of colesevelam on reducing common glycaemic parameters such as fasting blood glucose and oral glucose tolerance test (OGTT). Colesevelam treatment improved had no effect on body weight nor on food intake, suggesting its action is exerted through increased levels of active GLP-1. It was also effective at reducing β-cell islets degeneration when compared to the control. This article was classified as category II. This study is in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 1, the article has no information on housing nor husbandry, has information on sample size, has no information on sample size calculation, blinding, randomisation nor acclimatisation and has information on sex (all animals were male).

7.1.4 Dopamine-2 Receptor Agonist: Bromocriptine

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND bromocriptine' on 27/02/2017, yielding no results.

7.1.5 DPP-IV inhibitor: Sitagliptin (115,122,132–138)

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND sitagliptin' on 27/02/2017, yielding 12 results which were screened by title and abstract. Of these, ten (10) articles were included in this section.

Results: most studies show a positive effect on glycaemic control with sitagliptin significantly lowering plasma glucose, HbA1c and increasing insulin secretion. Two studies did not find any significant change in these outcomes. They also show an increase in aGLP-1 levels and a decrease in triglycerides. An increase in general β-cell function is also reported by the glucose/insulin ration and the HOMA-beta index. This is supported by no evidence of induction of pancreatitis and general improvement of islet histology (reduction of β-cell vacuolation, inflammation, fibrosis and apoptosis). Additionally, a reduction in the expression of IL-1β and genes involved in apoptosis induction was also reported. All articles were classified as category II. Eight (8) studies are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 10, 20% have information on housing; 80% on husbandry, 100% on sample size, none on sample size calculation nor blinding, 20% on randomisation; 40% on acclimatisation; and 90% on sex (all animals were male).

7.1.6 GLP-1 agonist: Exenatide (106,139–145)

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND exenatide' on 27/02/2017, yielding 14 results which were screened by title and abstract. Of these, eight (8) articles were included in this section.

Results: all studies show a positive effect of exenatide on blood glucose and/or HbA1c. Exenatide was also able to increase insulin levels after feeding, indicating increased β-cell activity. Additionally, it reduced triglyceride levels in female ZDF rats and improved islet histology (e.g. reduced β-cell vacuolation). All articles were classified as category II. All studies are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 8, 25% have information on housing; 88% on husbandry, 100% on sample size, none on sample size calculation nor blinding, 25% on randomisation; 50% on acclimatisation; and 88% on sex (6 used only male animals and one used both).

7.1.7 Metiglinide: Repaglinide

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND repaglinide' on 27/02/2017, yielding no results. The same string was used to search for studies replacing repaglinide by nateglinide, however, this search also returned no results.

7.1.8 PPAR-γ agonist: Rosiglitazone (105,107,119,126,146–180)

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND rosiglitazone' on 27/02/2017, yielding 47 results which were screened by title and abstract. Of these, thirty-nine (39) articles were included in this section.

Results: most studies show a significant reduction in glycaemic (plasma glucose, HbA1c, OGTT) and lipidemic parameters (LDL, free fatty acids (FFA), total cholesterol and triglycerides). Some studies in which no effect on glycaemic control was achieved were performed in older animals, in which β-cell degeneration is already at an advanced stage and this reason was cited as a possible cause for the lack of efficacy. Rosiglitazone reduced insulin levels during the hyperinsulinemic phase and increased them during the hypoinsulinemic phase. It also increased HOMA-IR and insulin sensitivity indexes, improved islet morphology (increased islet area and insulin content). Many studies showed a considerable reduction of food intake with an increase in body weight. Furthermore, rosiglitazone could reverse the metabolic phenotype of the diabetic heart, reduce the number of apoptotic cardiomyocytes and infarction size. It can normalise gene expression in the liver and decrease hepatic fat. It has also been showed to improve nephropathy by decreasing kidney weight, hypertrophy and urinary glucose. Although according to one study it can restore endothelium-dependent vasorelaxation, it had no effect on endothelium's mechanical properties. All articles but one were classified as category II. Of the 38 category II studies, thirty-two (32) are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 39, 46% have information on housing; 82% on husbandry, 90% on sample size, none on sample size calculation nor blinding, 41% on randomisation; 46% on acclimatisation; and 90% on sex (32 used only male animals, one used only female animals and 2 used both).

7.1.9 SGLT-2 inhibitor: Canagliflozin (133,181–186)

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND canagliflozin' on 27/02/2017, yielding 7 results which were screened by title and abstract. Of these, six (6) articles were included in this section.

Results: all studies show a positive effect of canagliflozin on plasma glucose, HbA1c and glucose AUC in OGTT tests. Some studies also report an increase in plasma insulin levels and aGLP-1, although there are conflicting results regarding aGLP-1. Canagliflozin was able to restore insulin immunoreactivity in ZDF rat islets back to similar levels as the lean controls. Furthermore, it reduced β-cell vacuolation and increase β-cell general activity measured by glucose/insulin ratio following an OGTT. All articles were classified as category II. All studies are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 6, 17% have information on housing; 83% on husbandry, 100% on sample size, none on sample size calculation, blinding nor randomisation; 50% on acclimatisation; and 100% on sex (all animals were male).

7.1.10 Sulphonylurea: Glibenclamide (111,187)

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND (glibenclamide OR glyburide)' on 27/02/2017, yielding six (6) results which were screened by title and abstract. Of these, two (2) articles were included in this section.

Results: one study showed a positive effect of glibenclamide on OGTT, lowering the glucose AUC_{0-4h} when compared to control. However, the other study showed no effect on blood glucose, insulin or HbA1c levels besides no improvement on islet histology. Both articles were classified as category II. Only one study is in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 2, 50% have information on housing, 100% on husbandry, 50% sample size, none on sample size calculation nor blinding, 50% on randomisation, none on acclimatisation; and 100% sex (all animal were male).

7.1.11 α-glucosidase Inhibitor: Acarbose (31,188–190)

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND acarbose' on 27/02/2017, yielding 4 results which were screened by title and abstract. All four articles were included in this section.

Results: all studies show a positive effect of acarbose on reducing blood glucose and HbA1c. Acarbose has also been shown to reduce basement membrane thickening with a modest effect on cell density in ZDF retinopathy. It can also reduce free fatty-acid levels in the blood as well as cholesterol. All articles were classified as category II. All studies are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 4, 25% have information on housing, 75% on husbandry, 75% on sample size, none on sample size calculation nor blinding, 25% on randomisation, 25% on acclimatisation; and 75% on sex (all animal were male).

7.2 Are ineffective drugs in humans also ineffective in this model?	Score
Of the six identified classes of ineffective drugs, only two have been tested in ZDF rats, none being in line with the clinical findings. The efficacy of the other five classes in this model remains unclear.	0.10

7.2.1 11-Beta hydroxysteroid dehydrogenase inhibitors

Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND ("11-beta hydroxysteroid dehydrogenase inhibitor" OR "11-beta hydroxysteroid dehydrogenase inhibitors" OR "11 beta-HSD" OR "11βHSD")' on 27/02/2017, yielding no results.

7.2.2 Adenosine A1 receptor agonists
Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND "Adenosine A1 receptor agonist" OR "Adenosine A1 receptor agonists" OR AA1RA' on 27/02/2017, yielding 1 result which was screened by title and abstract. This article was included in this section.
7.2.2.1 CVT-3619 (191)
Results: the only study in ZDF rats showed CVT-3619 reduced FFA levels and inhibited lipolysis. This article was classified as category I. Quality Assessment: according to the pre-specified criteria and n = 1, the article has information on housing, on husbandry, on sample size, no information on sample size calculation, blinding, randomisation nor acclimatisation; and has information on sex (all animals were male).
7.2.3 Nicotinic α-7 receptor agonists
Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND "nicotinic α-7 receptor" OR "Nicotinic α-7 receptor agonist" OR "α7NR"' on 27/02/2017, yielding no results.
7.2.4 TGR5 receptor agonists
Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND ("G protein-coupled bile acid receptor 1 agonist" OR GPBAR1) ("TGR5 receptor" OR Gpbar1 OR M-BAR OR GPR131 OR BG37 OR Axor109)' on 27/02/2017, yielding 1 result which was screened by title and abstract. This article was not included in this section.
7.2.5 Protein tyrosine phosphatase 1B inhibitors
Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND (PTP1B OR "protein tyrosine phosphatase 1B inhibitor" OR "protein tyrosine phosphatase 1B inhibitors")' on 27/02/2017, yielding 1 result which was screened by title and abstract. This article was not included in this section.
7.2.6 Fructose-1,6-bisphosphatase inhibitors
Methodology: studies were searched on PubMed with the string '("ZDF rat" OR "ZDF rats" OR "zucker diabetic fatty rat" OR "zucker diabetic fatty rats") AND ("fructose-1,6-bisphosphatase inhibitor" OR "fructose-1,6-bisphosphatase inhibitors" OR "FBPase inhibitor" OR "FBPase inhibitors")' on 27/02/2017, yielding 5 results which were screened by title and abstract. Of these, three (3) articles were included in this section.
7.2.6.1 CS-917 (192–194)
Results: all three studies showed a significant reduction in plasma glucose levels and inhibition of gluconeogenesis. All articles were classified as category II. None of these studies is in line with the clinical findings. **Quality Assessment**: according to the pre-specified criteria and n = 3, no articles have information on housing, 67% have information on husbandry, 100% on sample size, none on sample size calculation nor blinding, 33% on randomisation, 33% on acclimatisation; and 100% on sex (2 used only male animals and one used both sexes).

7.3 Have drugs with different mechanisms of action and acting on different pathways been tested in this model? If so, which?	Score
Yes, partially. **Remarks**: out of the 17 identified drug classes tested and/or used to treat type 2 diabetes, 10 were tested in ZDF rats.	1.18

8. ENDPOINT VALIDATION	
8.1 Are the endpoints used in preclinical studies the same or translatable to the clinical endpoints?	Score
Yes. **Remarks**: most studies performed in ZDF rats with agents to treat type 2 diabetes have used glycaemic parameters (e.g. glycaemia, HbA1c, OGTT) or other measures of insulin sensitivity. These measurements also often represent the primary outcomes of trials testing new drugs for the treatment of type 2 diabetes, which aim to control glycaemia.	8
8.2 Are the methods used to assess preclinical endpoints comparable to the ones used to assess related clinical endpoints?	Score
Yes. **Remarks**: the biochemical methods used in the preclinical studies are the same or similar to the ones used to measure glycaemic parameters in humans.	3

MODEL NAME	db/db mouse (BKS.Cg-Dock7^m +/+ Lepr *db*/J*)*
INDICATION	Type 2 Diabetes Mellitus (T2D)
VALIDATION DATE	10.07.2017
TOTAL SUBSECTIONS	59
TOTAL SCORE	67.36
UNCERTAINTY FACTOR (%)	23.7

HISTORICAL BACKGROUND
In 1966 at Jackson Laboratories, a mutation in the inbred mouse strain C57BL/Ks occurred which caused a metabolic syndrome similar to the human diabetes mellitus (195). The mice manifested hyperglycaemia, hyper- then hypoinsulinemia and histological changes in the pancreatic islets. This is an autosomal recessive mutation and the strain was called db/db mouse (recessive diabetic) in a similar fashion to ob/ob (recessive obese) mouse (195). Both strains have mutations in the leptin pathway, the db/db affecting the leptin receptor while the ob/ob affects leptin itself. This nomenclature replaced the old 'Ob-protein and Ob-receptor' used before to differentiate both strains. The misty mutation (*Dock7^m*) was introduced to maintain the diabetes gene as homozygotes for the Lepr^{db} mutation are infertile besides helping in the identification of homo- and heterozygotes by colour changes in the fur (196). The db/db mouse is commercialised with two different backgrounds: C57BL/6 (resistant to the development of diabetes) and C57BL/Ks (susceptible), which also known as C57BL/KsJ (196–200).

1. EPIDEMIOLOGICAL VALIDATION	
1.1 Is the model able to simulate the disease in the relevant sexes?	Score
Yes, completely. **Remarks**: both male and female db/db mice can develop diabetes. The diabetic condition is somewhat more pronounced in male than in female mice (195,201). This specific female resistance to the development of diabetes is in line with what is seen in humans, due to a possible protective effect of oestrogen on pancreatic beta-cells (7,8).	2
1.2 Is the model able to simulate the disease in the relevant age groups (juvenile, adult or ageing)?	Score
Yes, partially. **Remarks**: Although the time to onset of diabetes has been decreasing in the past years, the db/db mouse still develops diabetes somewhat earlier than humans (9,196). However, it progresses into adulthood and ageing phases, similarly to the human disease (202–206).	1

2. SYMPTOMATOLOGY AND NATURAL HISTORY (SNH) VALIDATION	
2.1 Is the model able to replicate the symptoms commonly present in this disease? If so, which ones?	Score
Yes, completely. The most common symptoms of diabetes type 2 are polyphagia, polydipsia, polyuria, weight loss, fatigue, healing impairment and obesity (15). The latter may not be considered a symptom per se, is better classified as a co-morbidity. However, due to the high prevalence of this condition in T2D patients and likely association to the pathophysiology, it was also included in this section. Most of these symptoms are directly connected to the glycaemia and dyslipidaemia and therefore they were translated into the following parameters: hyperglycaemia, hyperinsulinemia followed by hypoinsulinemia, dyslipidaemia (divided into hypercholesterolemia and hypertriglyceridemia according to the common human pathological profile) and obesity.	13
2.1.1 Symptoms modelled	
Hyperglycaemia: db/db mice have increased blood glucose levels of up to 500% normal levels (195,197–199,202,203,205,207,208). **Hyperinsulinemia followed by hypoinsulinemia**: insulin levels are increased in db/db mice as early as 10 days of age. Over time, with extenuation of β-cells, insulin levels drop to normal or below normal (197,200–203,205,208,209). **Hypercholesterolemia**: cholesterol levels are significantly increased in db/db mice starting at 5 weeks of age (206,210). **Hypertriglyceridemia**: triglycerides levels are higher in db/db mice than in controls of the same background, starting with a moderate increase at 5 weeks and progressively increasing with age up to 19 weeks (206,210). **Obesity**: db/db mice are significantly heavier than their lean littermates at a young age, but tend to significantly lose weight after overt diabetes, which is consistent with the sudden weight loss sometimes seen in humans (195,199,202,203,208).	
2.2 Is the natural history of the disease similar to humans regarding:	
2.2.1 Time to onset;	Score
No. **Remarks**: type 2 diabetes is usually diagnosed around 40 years, although age at diagnosis has been decreasing (9). The onset of diabetes in db/db mice may vary from 4 to 8 weeks (196). This happens at quite an early stage of life when compared to humans: mice 4 weeks of age are in a similar stage of development as humans between 6 months and 10 years old while mice at 8 weeks correspond to a human at around 20 years (211,212).	0
2.2.2 Disease progression;	Score
Yes, partially. **Remarks**: Diabetes progression in db/db mice follows much of the human disease. The increase in sugar/fat intake in diets leads to an increase in insulin secretion, reducing the ability to activate the TK receptor (between 4 and 8 weeks of age). Consequently, this then leads to an increase in glycaemia and β-cell degeneration, finally leading to β-cell apoptosis and glucose intolerance (202–206). The whole process, however, happens at quite an early stage of life when compared to humans: mice 4 weeks of age are in a similar stage of development as humans between 6 months and 10 years old while mice at 8 weeks correspond to a human at around 20 years (211,212). Although as mentioned before, the age of onset of diabetes has been decreasing in the past years (9).	1

As for common diabetic complications, such as nephropathy, retinopathy and neuropathy, db/db mice can partially model them. They are able to develop renal lesions (such as albuminuria, glomerular basement thickening, loss of podocytes and moderate mesangial matrix expansion) but fail to reproduce mesangiolysis, nodular mesangial sclerosis and severe tubulointerstitial fibrosis, not developing progressive renal insufficiency (213–218). Regarding retinopathy, some characteristics such as pericyte loss, degeneration of the blood-retinal barrier (BRB), apoptosis of neuro-retinal cells, thickening of capillary basement membranes, glial reactivation and vascular proliferation are present while it cannot model retinal neovascularization (219–224). Likewise, db/db mice can partially develop neurological complications, such as decreased motor nerve conduction velocity (MNCV), the absence of large myelinated fibres, axonal atrophy and dystrophy (myelinated and unmyelinated fibres); and degeneration of myelin sheath, but do not develop neurotic dystrophy (225–229).	1

2.2.3 Duration of symptoms;	Score
Yes, completely. **Remarks**: like humans, once insulin resistance and glucose intolerance are set, β-cell apoptosis follows, worsening the diabetic condition progressively and leading to overt diabetes and diabetic complications (195,200,202).	2

2.2.4 Severity.	Score
Yes, partially. **Remarks**: the biochemical parameters in db/db mice are similar to humans in severity, is considerably increased. However, the complications that follow after the onset of overt diabetes are somewhat milder in db/db mice (e.g. lack of mesangiolysis and tubulointerstitial fibrosis in nephropathy; neovascularisation in retinopathy or neurotic dystrophy in neuropathy).	1

3. GENETIC VALIDATION	
3.1 Does this species also have orthologous genes and/or proteins involved in the human disease? If so, which?	Score
Yes, completely. Although is widely accepted that diabetes type 2 has genetic factors involved, reliable tracking and identification of such genes remain a challenge (35–42). This is often referred to as the 'missing heritability' (36). In some cases, T2D can be caused by a single gene (such as mutations on HNF4A for Maturity-Onset-Diabetes of the Young type 1 – MODY1 – or PTF1A for neonatal diabetes) (35,37). Nonetheless, in most cases it is a polygenic and heterogeneous disease, meaning multiple genes might be involved and different combinations of polymorphisms can lead to the pathological phenotype, which makes the genetic validation of animal models at best extremely challenging (35,37,39–42). With the advent of Genome-Wide Association Studies (GWAS), many genes have been associated with an increased risk of developing type 2 diabetes, shedding light on the complex genetic architecture of the disease (35,43,44). However, three of them are better established as likely genetic factors: TCF7L2, KCNJ11 and PPARG (35,38,45). Therefore, these were the only ones included in this section. All three genes are present in db/db mice (230–232).	8

3.1.1 TCF7L2
TCF7L2 is a gene (also known as also known as TCF-4 or β-catenin interacting protein) located at 10q25.2-q25.3 in humans and at 1q55 in rats that codes a transcription factor involved in the WNT signalling pathway (46,49–52). This is the strongest and most well-replicated genetic factor associated to T2D (35,38,43,44,53,54). It acts as a nuclear receptor for β-catenin, which is involved in the secretion of GLP-1 in gut endocrine cells and various other genes (50,51,55). There are also reports that indicate a possible role in the incretin axis, adipocyte function and glucose production by the liver, despite some of the results being conflicting (52,56–60).
3.1.2 KCNJ11
KCNJ11 is a gene (also known as KIR6.2) located at 11p15.1 in humans and at 1q22 in rats that has been associated with an increased risk of developing T2D (43–45,48,61). It codes for a major subunit of the ATP-sensitive K^+ channel, an inward-rectifier potassium ion channel present in the pancreatic islets with direct influence over insulin secretion (62,63). Mutations in this gene are connected to neonatal diabetes (35).
3.1.3 PPARG
PPARG is a gene that codes for the transcription factor peroxisome proliferator-activated receptor-γ (PPARγ) located at 3p25.2 and at 4q42 in rats in humans (47,64). The PPARG is involved in adipocyte regulation, fat accumulation and glucose metabolism and it is stimulated by insulin in a ligand-dependent manner (65–68). Two isoforms exist in humans: PPARγ1 and PPARγ2, the first being expressed in most tissues and the second one mostly in the liver and adipose tissue (69). The isoform 2 is also the target of thiazolidinediones, such as rosiglitazone.

3.2 If so, are the relevant genetic mutations or alterations also present in the orthologous genes/proteins?	Score
Unclear.	0.5

3.2.1 TCF7L2
Unclear.
In humans, the rs12255372 and rs7903146 alleles are the most strongly associated variants with carriers of such type having reduced insulin secretion but not increased insulin resistance (70). Being homozygous for the high-risk allele doubles the chance of developing T2D (35). No studies investigating specific genetic alterations in this gene could be found in PubMed in db/db mice using the string ("db/db mouse" OR " db/db mice") AND TCF7L2 on 05/07/2017.
3.2.2 KCNJ11
Unclear.
In humans, the commonest polymorphism is a glutamate to lysine substitution at position 23 (E23K or rs5219) (71–74). This mutation is linked to a reduction in channel sensitivity, increasing the signal threshold for the release of insulin and impairing serum insulin response (62,71,72,75). Homozygous carriers of this mutation (KK) have an almost double risk of developing T2D when compared to non-carriers (73).
No studies investigating specific genetic alterations in this gene could be found in PubMed in db/db mice using the string ("db/db mouse" OR " db/db mice") AND (KCNJ11 OR KIR6.2 OR "KIR 6.2") on 05/07/2017.

3.2.3 PPARG

Unclear.

In humans, the SNP rs1801282 consists of a proline substitution for alanine in position 12 (Pro12Ala or P12A) of PPARγ2 (the other isoform is not affected), the proline allele is associated with an increased risk of developing T2D, higher BMI and decreased insulin sensitivity while the alanine one confers resistance (43–45,73,74,76–78). Even though the mechanism is still unclear, the Ala variant has a lower transactivation

Efficiency, reducing stimulation of PPARG target genes and therefore also reducing the levels of adipose tissue mass accumulation (76). Nonetheless, there are also reports of increased risk of cardiovascular disease in carriers of this polymorphism (79,80).

No studies investigating specific genetic alterations in this gene could be found in db/db mice in PubMed using the string ("db/db mouse" OR " db/db mice") AND PPARg on 05/07/2017.

3.3 If so, is the expression of such orthologous genes and/or proteins similar to the human condition?	Score
Unclear.	0.2

3.3.1 TCF7L2

Unclear.

Two studies found the TCF7L2 gene is overexpressed in carriers of the risk genotype in pancreatic islets of T2D patients, but no changes in TCF7L2 expression were found by study of the Diabetes Genome Anatomy Project (DGAP) (35,56,81). One study by Shu and colleagues showed a possible protective effect on the islets against glucose- and cytokine-induced apoptosis and function impairment. This study was later retracted due to image duplication and concerns over data reliability (82). Two studies analysed expression of TCF7L2 in db/db mice in two different tissues: fat tissue and islets. TCF7L2 expression was increased in fat tissue of db/db mice, similarly to what is reported in humans while in islets, it is decreased (57,233,234).

3.3.2 KCNJ11

Unclear.

KCNJ11 is underexpressed in human islets, which is consistent with impaired β-cell insulin release (86,87). Nevertheless, no studies investigating expression levels of KCNJ11 could be found in db/db mice using the string ("db/db mouse" OR " db/db mice") AND (KCNJ11 OR KIR6.2 OR "KIR 6.2") on 07/07/2017.

3.3.3 PPARG

Unclear.

PPARG is overexpressed in adipose tissue and underexpressed in the liver of diabetic humans (35). In db/db mice, during endotoxemia PPARG is underexpressed in White Adipose Tissue (WAT) (235). A knockout mouse of Wdr13 with C57BL6 or db/db background significantly increases expression of PPARG in fat tissue (236). A study of gene expression variation in db/db mice during the day showed significantly increased PPARG expression levels in the aorta during certain periods but did not analyse adipose tissue, islets or liver (237). No studies specifically assessing PPARG expression in adipose tissue and liver could be found using the string ("db/db mouse" OR " db/db mice") AND PPARG on 07/07/2017.

4. BIOCHEMICAL VALIDATION	
4.1 If there are known pharmacodynamic (PD) biomarkers related to the pathophysiology of the disease, are they also present in the model?	Score
Yes, completely. **Remarks**: Glycaemic markers (blood glucose and HbA1c), cholesterol and triglycerides levels are directly related to the development of diabetes and diabetic complications, being used as PD biomarkers.	3
4.1.1 Glycaemic markers (blood glucose, HbA1c)	
All glycaemic markers can be measured in db/db mice.	
4.1.2 Cholesterol	
Cholesterol levels can be measured in db/db mice.	
4.1.3 Triglycerides	
Triglycerides levels can be measured in db/db mice.	
4.2 Do these PD biomarkers behave similarly to humans'?	Score
Yes, completely.	2
4.2.1 Glycaemic markers (blood glucose, HbA1c)	
All glycaemic markers are increased in db/db mice (see section 2.1.1).	
4.2.2 Cholesterol	
Cholesterol levels are increased in db/db mice (see section 2.1.1).	
4.2.3 Triglycerides	
Triglycerides levels are increased in db/db mice (see section 2.1.1).	
4.3 If there are known prognostic biomarkers related to the pathophysiology of the disease, are they also present in the model?	Score
Yes, completely. **Remarks**: Glycaemic markers are also prognostic markers as higher glycaemic levels can potentially lead to faster worsening of the diabetic condition (88–90). The same for cholesterol and triglycerides, which are directly involved in cardiovascular damage and diabetic complications (91,92).	3
4.3.1 Glycaemic markers (blood glucose, HbA1c)	
All glycaemic markers can be measured in db/db mice.	
4.3.2 Cholesterol	
Cholesterol levels can be measured in db/db mice.	
4.3.3 Triglycerides	
Triglycerides levels can be measured in db/db mice.	
4.4 Do these prognostic biomarkers behave similarly to humans'?	Score
Yes, completely. **Remarks**: Like humans, db/db mice have increased levels of blood glucose, cholesterol and triglycerides.	2

4.4.1 Glycaemic markers (blood glucose, HbA1c)
All glycaemic markers are increased in db/db mice (see section 2.1.1).
4.4.2 Cholesterol
Cholesterol levels are increased in db/db mice (see section 2.1.1).
4.4.3 Triglycerides
Triglycerides levels are increased in db/db mice (see question 2.1.1).

5. AETIOLOGICAL VALIDATION	
5.1 Is the aetiology of the disease similar to humans'?	Score
Yes, partially. **Remarks**: diabetes type 2 is strongly associated with obesity, high-fat/carbs diet and sedentary lifestyle (9,93). Lifestyle changes such as a healthful diet, BMI control and reduction of smoking/alcohol intake could prevent as much as 90% of type diabetes cases (93,94). Diabetes is caused in the db/db mouse by a point mutation (Gly → Thr) in the OB-R (leptin receptor) sequence located at chromosome 4 (238,239). The lack of signalling promotes the hypersecretion of circulating leptin, but the defective receptor prevents it from properly regulating the size of the body fat depot (34). This was first suggested in several parabiosis experiments in which controls and diabetic mice (ob and db) had their circulatory systems connected. Ob/ob mice were able to lose weight and reduce food intake, indicating the mutation interfered with the action of a circulating factor (leptin) while db/db mice did not have any improvement, indicating its mutation affected the signal reception (leptin receptor) (207,240–242). This hypothesis is further supported by the lack of response of db/db mouse to recombinant leptin (243–246). In humans, although leptin has a similar function, it seems to be less critical for the regulation of energy expenditure as inactivating mutations in its receptor have milder effects than in rodents (97). Several genome-wide association studies (GWAS) have conflicting results regarding the association of polymorphisms in the Lepr gene and risk of diabetes (98–100). Nevertheless, the existence of genetic factors is widely accepted, although they probably result from the product of several small to moderate gene effects (35–42). For overt diabetes to develop in humans, it is necessary to have a genetic predisposition, β-cell dysfunction and insulin resistance (usually acquired due to obesity). In that sense, db/db mice can only partially reproduce this dysfunction as diet restriction only slightly improves their condition (208). Male db/db mice also have a somewhat more severe manifestation of diabetes than female mice. This fact is also in line with the specific female resistance to the development of diabetes seen in humans, due to a possible protective effect of oestrogen on pancreatic beta-cells (7,8).	7

6. HISTOLOGICAL VALIDATION	
6.1 Do the histopathological structures in relevant tissues resemble the ones found in humans?	Score
Yes, partially. There is some evidence of β-cell mass reduction and pathogenic role of amyloid deposits in human patients, even though conflicting results to these notions are also present (101–104). Nonetheless, they were included to better characterise the models in this sheet. The pancreas was the only tissue selected for the histological validation because it primarily affects the pancreatic islets, whose long-term damage then leads to diabetic complications.	5.5

6.1.1 Histopathological features modelled
B-cell mass: as seen in diabetic patients, the β-cell mass in db/db mice islets is severely diminished with the development of overt diabetes (195,203–205).
6.1.2 Histopathological features partially modelled
Islet morphology: islets become irregularly delineated with the advancement of diabetes, showing signs of hypertrophy, mixtures of acinar and islet cells and partial or total degranulation and necrosis of β-cells (195,198,203–205).
6.1.3 Histopathological features not modelled
Amyloidosis: islet amyloid polypeptide (IAPP) or amylin start increasing at 8 weeks but decrease significantly at 24 weeks. Nonetheless, there is no report of formation of the characteristic amyloid plaques as seen in humans (247).

7. PHARMACOLOGICAL VALIDATION	
7.1 Are effective drugs in humans also effective in this model?	Score
Yes, partially. **Remarks**: one drug from each class approved by FDA and/or EMA for the treatment of type 2 diabetes was included. Studies that used db/db mice with C57BL/6 background (also known as B6.BKS(D)-Lepr db/J, #00697 at Jackson Laboratories) were excluded. In consultation with the main provider of this model, Jackson Laboratories, the following assumptions/decisions were made: studies that used db/db mice with C57BLK6/J or C57BLK/6 background were considered to have used the C57BL/6 background and therefore were also excluded. Studies with mice of C57BL/KsOlaHsd-Lepr and C57BLKS/J lar m+/Lepr^{db} were also excluded as it is likely that a mutation was engineered into the strain. A C57BL/KFJ background was assumed to be a C57BL/KsJ background (likely typo) as the references cited referred to this strain. Additionally, studies that did not disclose which background strain was used (i.e. just mentioned db/db mice) were included but were flagged with a 'missing background' note. If a study disclosed the control's strain but not specifically the db/db mouse's, it was assumed the diabetic mice had the same background as the control.	3.88
7.1.1 Amylin Agonist: Pramlintide	
Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND pramlintide' on 10/07/2017, yielding no results.	
7.1.2 Biguanide: Metformin (248–281)	
Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND metformin' on 10/07/2017, yielding 68 results which were screened by title and abstract. A study retrieved through the search string for rosiglitazone with a metformin monotherapy arm was also screened. Of these, thirty-four (34) articles were included in this section. **Results**: most studies show a reduction of glycaemia measures, such as plasma glucose, HbA1c, OGTT upon administration of metformin. Frequently, metformin has been shown to reduce plasma insulin levels during the hyperinsulinemic phase, although such benefit was not reported consistently. Some studies also report an improvement on the lipidic profile (reduction of LDL and FFA; and increase of HDL) and reduction of body weight and food intake. However, these results are often conflicting with other studies showing no effect at all on these parameters. A reduction of HOMA-IR is also often reported, including improvement on islet histopathology (e.g. reduced loss of islet boundaries and vacuolar degeneration). In the few studies which investigated metformin's effect on diabetic retino- and nephropathy, it failed to show any significant improvement.	

Two (2) articles were classified as category I and thirty-two (32) as category II. Of the latter, twenty-eight (28) studies are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 34, 29% have information on housing, 79% on husbandry, 94% on sample size, none on sample size calculation nor blinding, 44% on randomisation, 41% on acclimatisation, 88% on sex (31 used only male animals and 3 only female animals); and 76% on background strain.

7.1.3 Bile Acid Sequestrant: Colesevelam

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND cholestyramine' on 10/07/2017, yielding 2 results which were screened by title and abstract. None of these articles was included in this section.

7.1.4 Dopamine-2 Receptor Agonist: Bromocriptine

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND bromocriptine' on 10/07/2017, yielding 1 result. This article was not included in this section.

7.1.5 DPP-IV inhibitor: Sitagliptin (282)

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND sitagliptin' on 10/07/2017, yielding 5 results which were screened by title and abstract. Of these, one (1) article was included in this section.

Results: the only study included reports a significant reduction of blood glucose, triglycerides and LDL while an increase in HDL. Both glucose tolerance (OGTT) and HOMA index were improved after treatment with sitagliptin. This article was classified as category II. This study is in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 1, this article has no information on housing, husbandry, sample size, sample size calculation, blinding, randomisation, acclimatisation nor sex; while it has information on background strain.

7.1.6 GLP-1 agonist: Exenatide (283–302)

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND exenatide' on 10/07/2017, yielding 60 results which were screened by title and abstract. Of these, twenty (20) articles were included in this section.

Results: most studies report a positive effect of exenatide on glycaemic parameters (blood glucose, HbA1c) and glucose tolerance. Administration of exenatide has also been shown to increase plasma insulin levels, improve the HOMA-IR and islet histology, increase β-cell mass and the count of insulin-positive cells. Additionally, it improved diabetic nephropathy (e.g. by reducing glomerular hypertrophy and mesangial matrix expansion) despite no evident effect on blood glucose. Some studies have also shown an improvement in the overall lipidic profile by reducing in free-fatty acids (FFA) and LDL levels while increasing HDL. Three (3) articles were classified as category I and seventeen (17) as category II. Of the latter, twelve (12) studies are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 20, 30% have information on housing, 65% on husbandry, 100% on sample size, none on sample size calculation nor blinding, 20% on randomisation, 10% on acclimatisation, 80% on sex (12 used only male animals and 4 only female animals); and 85% on background strain.

7.1.7 Metiglinide: Nateglinide (303)

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND repaglinide' on 10/07/2017, yielding no results. The same string was used to search for studies replacing repaglinide by nateglinide, '("db/db mouse" OR "db/db mice") AND nateglinide'), yielding 1 result. This article was included in this section.

Results: the only study did not show any significant effect on blood glucose, plasma nor pancreatic insulin. This article was classified as category II. This study is not in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 1, this article has no information on housing nor husbandry, has information on sample size, none on sample size calculation, blinding, randomisation nor acclimatisation, has information on sex (all animal were male) and none on background strain was.

7.1.8 PPAR-γ agonist: Rosiglitazone (163,179,253,266,269,281,304−392)

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND rosiglitazone' on 10/07/2017, yielding 128 results which were screened by title and abstract. Of these, ninety-six (96) articles were included in this section. One likely relevant article was not included because it was not available via university's library:

Nanayakkara et al, Current Pharmaceutical Design. 2013, 19(27):4839-4847.

Results: most studies show a significant reduction in glycaemic (plasma glucose, HbA1c, OGTT) parameters. Although the majority of studies report an improvement in lipidemic parameters (LDL, free fatty acids (FFA), total cholesterol and triglycerides), some show conflicting results with an increase in total cholesterol and LDL and a decrease in HDL. Rosiglitazone reduced and increased insulin levels during the hyperinsulinemic and hypoinsulinemic phases, respectively. It also improved HOMA-IR and insulin sensitivity indexes, including islet morphology (increased β-cell area and number of insulin-positive cells). Many studies showed a considerable increase in body weight and body weight gain. Rosiglitazone treatment has been shown to improve fibrosis and inflammatory infiltration in the liver. Additionally, some studies show a deleterious effect on cardiac gene expression, mostly by activation of pro-apoptotic genes, which is likely related to rosiglitazone's cardiotoxicity. Sixteen (16) articles were classified as category I and eighty (80) as category II. Of the latter, seventy-eight (78) studies are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 96, 36% have information on housing, 82% on husbandry, 92% on sample size, none on sample size calculation, 1% on blinding, 35% on randomisation, 38% on acclimatisation, 91% on sex (75 used only male animals, 10 only female animals and 2 both sexes); and 75% on background strain.

7.1.9 SGLT-2 inhibitor: Canagliflozin (184)

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND canagliflozin' on 10/07/2017, yielding 1 result which was screened by title and abstract. This article was included in this section.

Results: the only study included shows a significant decrease in blood glucose of db/db mice treated with canagliflozin. No other parameters were analysed in this study in db/db mice. This article was classified as category II. This study in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 1, the article has information on sex (all animals were male) and background strain while no information on housing, husbandry, sample size, sample size calculation, blinding, randomisation nor acclimatisation was present.

7.1.10 Sulphonylurea: Glibenclamide (248,249,303,393–395)

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND (glibenclamide OR glybenclamide)' on 10/07/2017, yielding nine (9) results which were screened by title and abstract. A study retrieved through the search string for acarbose with a glibenclamide monotherapy arm was also screened. Of these, seven (7) articles were included in this section.

Results: With a thin margin, most studies show a positive effect of glibenclamide treatment on plasma glucose and insulin. One study also reports benefits for the integrity of the blood-brain barrier including a reduction in influx and activation of apoptosis-related biomarkers such as caspase-3 while another reports an increase in irisin release and reduction of triglycerides levels. Three (3) studies do not report any effect of glibenclamide on blood glucose, HbA1C and/(n)or plasma insulin levels. All articles were classified as category II. Four (4) studies are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 7, 43% have information on housing, 57% on husbandry, 100% on sample size, none on sample size calculation nor blinding, 14% on randomisation, none on acclimatisation, 100% on sex (6 used only male animals and 1 used only female animals); and 57% on background strain.

7.1.11 α-glucosidase Inhibitor: Acarbose (250,304,396–402)

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND acarbose' on 10/07/2017, yielding 15 results which were screened by title and abstract. Of these, nine (9) articles were included in this section.

Results: most studies show a positive effect of acarbose in reducing blood glucose and/or HbA1c. One study reported no effect on blood glucose despite reductions in HbA1c and urinary glucose excretion; and amelioration of diabetic nephropathy. Conflicting results on body weight and food intake reduction are reported. Acarbose has been shown to improve wound healing and angiogenesis. Moreover, it significantly decreases sucrase-isomaltase (SI) complex expression. All articles were classified as category II. All studies but one are in line with clinical findings.

Quality assessment: according to the pre-specified criteria and n = 9, 67% have information on housing, 89% on husbandry, 100% on sample size, none on sample size calculation nor blinding, 44% on randomisation, 56% on acclimatisation, 100% on sex (8 used only male animals and one both sexes); and 100% on background strain.

7.2 Are ineffective drugs in humans also ineffective in this model?	Score
The efficacy of the six identified drug classes in this model remains unclear.	0.1

7.2.1 11-Beta hydroxysteroid dehydrogenase inhibitors

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND ("11-beta hydroxysteroid dehydrogenase inhibitor" OR "11-beta hydroxysteroid dehydrogenase inhibitors" OR "11 beta-HSD" OR "11βHSD")' on 10/07/2017, yielding no results.

7.2.2 Adenosine A1 receptor agonists: GS-9667/CVT-3619

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND ("Adenosine A1 receptor agonist" OR "Adenosine A1 receptor agonists" OR AA1RA)' on 10/07/2017, yielding no results.

7.2.3 Nicotinic α-7 receptor agonists

Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND ("nicotinic α-7 receptor" OR "Nicotinic α-7 receptor agonist" OR "α7NR")' on 10/07/2017, yielding no results.

7.2.4 TGR5 receptor agonists
Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND ("TGR5 receptor" OR Gpbar1 OR M-BAR OR GPR131 OR BG37 OR Axor109)' on 10/07/2017, yielding 5 results which were screened by title and abstract. No articles were included in this section.
7.2.5 Protein tyrosine phosphatase 1B inhibitor
Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND ("PTP 112" OR PTP112 OR PTP-112 OR ertiprotafib)' on 10/07/2017, yielding 19 results which were screened by title and abstract. No articles were included in this section.
7.2.6 Fructose bisphosphatase inhibitor/Gluconeogenesis inhibitor
Methodology: studies were searched on PubMed with the string '("db/db mouse" OR "db/db mice") AND ("fructose-1,6-bisphosphatase inhibitor" OR "fructose-1,6-bisphosphatase inhibitors" OR "FBPase inhibitor" OR "FBPase inhibitors")' on 10/07/2017, yielding 3 results which were screened by title and abstract. No articles were included in this section.

7.3 Have drugs with different mechanisms of action and acting on different pathways been tested in this model? If so, which?	Score
Yes, partially. **Remarks**: out of the 17 identified drug classes tested and/or used to treat type 2 diabetes, 9 were tested in db/db mice.	1.18

8. ENDPOINT VALIDATION	
8.1 Are the endpoints used in preclinical studies the same or translatable to the clinical endpoints?	Score
Yes. **Remarks**: most studies performed in db/db mice with agents to treat type 2 diabetes have used glycaemic parameters (e.g. glycaemia, HbA1c, OGTT) or other measures of insulin sensitivity. These measurements also often represent the primary outcomes of trials testing new drugs for the treatment of type 2 diabetes, which aim to control glycaemia.	8
8.2 Are the methods used to assess preclinical endpoints comparable to the ones used to assess related clinical endpoints?	Score
Yes. Remarks: the biochemical methods used in the preclinical studies are the same or similar to the ones used to measure glycaemic parameters in humans.	3

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

References

1.   Zucker LM, Zucker TF. Fatty, a new mutation in the rat. Journal of Heredity. 1961;52(6):275–8.
2.   King AJF. The use of animal models in diabetes research. British Journal of Pharmacology. 2012;166(3):877–94.
3.   Charles River Laboratory. ZDF Rat: ZDF-Leprfa/Crl [Internet]. [cited 2016 Nov 14]. Available from: http://www.criver.com/products-services/basic-research/find-a-model/zucker-diabetic-fatty-(zdf)-rat
4.   Srinivasan K, Ramarao P. Animal models in type 2 diabetes research: an overview. The Indian journal of medical research. 2007;125(3):451–72.
5.   Mulder GB, Luo S, Gramlich P. The Zucker Diabetic Fatty (ZDF) Rat: Diet Evaluation Study for the Induction of Type 2 Diabetes in Obese Female ZDF Rats [Internet]. Frontiers: A Journal of Women Studies. 2010 [cited 2016 Nov 21]. p. 1–4. Available from: http://www.criver.com/files/pdfs/rms/zdf/rm_rm_r_zdf_diet_eval_tech_sheet.aspx
6.   Corsetti JP, Sparks JD, Peterson RG, Smith RL, Sparks CE. Effect of dietary fat on the development of non-insulin dependent diabetes mellitus in obese Zucker diabetic fatty male and female rats. Atherosclerosis. 2000;148(2):231–41.
7.   Liu S, Mauvais-Jarvis F. Minireview: Estrogenic protection of β-cell failure in metabolic diseases. Endocrinology. 2010;151(3):859–64.
8.   Sowers JR, C A, RF G, DJ B, JR S, E B-C, et al. Diabetes Mellitus and Cardiovascular Disease in Women. Archives of Internal Medicine. 1998 Mar 23;158(6):617.
9.   King A, Bowe J. Animal models for diabetes: Understanding the pathogenesis and finding new treatments. Biochemical Pharmacology. 2016;99:1–10.
10. Katsuda Y, Ohta T, Miyajima K. Diabetic complications in obese type 2 diabetic rat models. Experimental …. 2014;63(October 2013):121–32.
11. Torres TP, Catlin RL, Chan R, Fujimoto Y, Sasaki N, Printz RL, et al. Restoration of hepatic glucokinase expression corrects hepatic glucose flux and normalizes plasma glucose in zucker diabetic fatty rats. Diabetes. 2009;58(1):78–86.
12. Paulsen SJ, Vrang N, Larsen LK, Larsen PJ, Jelsing J. Stereological assessment of pancreatic beta-cell mass development in male Zucker Diabetic Fatty (ZDF) rats: Correlation with pancreatic beta-cell function. Journal of Anatomy. 2010;217(5):624–30.
13. Hayek A, Woodside W. Correlation between morphology and function in isolated islets of the Zucker rat. Diabetes. 1979;28(6):565–9.
14. Etgen GJ, Oldham B a. Profiling of Zucker diabetic fatty rats in their progression to the overt diabetic state. Metabolism: clinical and experimental. 2000;49(5):684–8.
15. Vijan S. Type 2 diabetes. Annals of Internal Medicine. 2015;162(5):ITC1-16.
16. Pick A, Clark J, Kubstrup C, Levisetti M, Pugh W, Bonner-Weir S, et al. Role of apoptosis in failure of beta-cell mass compensation for insulin resistance and beta-cell defects in the male Zucker diabetic fatty rat. Diabetes. 1998;47(3):358–64.
17. Shibata T, Takeuchi S, Yokota S, Kakimoto K, Yonemori F, Wakitani K. Effects of peroxisome proliferator-activated receptor-alpha and -gamma agonist, JTT-501, on diabetic complications in Zucker diabetic fatty rats. British journal of pharmacology. 2000;130(3):495–504.
18. Belin De Chantemèle EJ, Vessières E, Guihot AL, Toutain B, Maquignau M, Loufrani L, et al. Type 2 diabetes severely impairs structural and functional adaptation of rat resistance arteries to chronic changes in blood flow. Cardiovascular Research. 2009;81(4):788–96.
19. Leonard BL, Watson RN, Loomes KM, Phillips ARJ, Cooper GJ. Insulin resistance in the Zucker diabetic fatty rat: A metabolic characterisation of obese and lean phenotypes. Acta Diabetologica. 2005;42(4):162–70.
20. Howarth FC, Al Kitbi MK, Hameed RS, Adeghate E. Pancreatic peptides in young and elderly Zucker type 2 diabetic fatty rats. Journal of the Pancreas. 2011;12(6):567–73.

21. Peterson RG, Shaw WN, Neel M, Little L a, Eichberg J. Zucker Diabetic Fatty Rat as a Model for Non-insulin-dependent Diabetes Mellitus. ILAR News. 1990;32(3):16–9.

22. Clark JB, Palmer CJ, Shaw WN. The diabetic Zucker fatty rat. Proc Soc Exp Biol Med. 1983;173(1):68–75.

23. Tokuyama Y, Sturis J, DePaoli a M, Takeda J, Stoffel M, Tang J, et al. Evolution of beta-cell dysfunction in the male Zucker diabetic fatty rat. Diabetes. 1995;44(12):1447–57.

24. Schmidt RE, Dorsey DA, Beaudet LN, Peterson RG. Analysis of the Zucker Diabetic Fatty (ZDF) type 2 diabetic rat model suggests a neurotrophic role for insulin/IGF-I in diabetic autonomic neuropathy. Am J Pathol. 2003;163(1):21–8.

25. Adami ANDREOLLO N, Freitas dos SANTOS E, Rachel ARAÚJO M, Roberto LOPES L, Adami Andreollo N. Rat's age versus human's age: what is the relationship? ABCD Arq Bras Cir Dig. 2012;25(1):49–51.

26. Vora JP, Zimsen SM, Houghton DC, Anderson S. Evolution of Metabolic and Renal Changes in the ZDF/Drt-fa Rat Model of Type II Diabetes. Journal of American Society of Nephrology. 1996;7(1):113–7.

27. Hoshi S, Shu Y, Yoshida F, Inagaki T, Sonoda J, Watanabe T, et al. Podocyte injury promotes progressive nephropathy in zucker diabetic fatty rats. Laboratory investigation; a journal of technical methods and pathology. 2002;82(1):25–35.

28. Chander PN. Nephropathy in Zucker Diabetic Fat Rat Is Associated with Oxidative and Nitrosative Stress: Prevention by Chronic Therapy with a Peroxynitrite Scavenger Ebselen. Journal of the American Society of Nephrology. 2004;15(9):2391–403.

29. Danis RP, Yang Y. Microvascular retionopathy in the Zucker diabetic fatty rat. Investigative Ophthalmology & Visual Science. 1993;34(7):2367–71.

30. Kim J, Kim CS, Sohn E, Kim H, Jeong IH, Kim JS. KIOM-79 Prevents Lens Epithelial Cell Apoptosis and Lens Opacification in Zucker Diabetic Fatty Rats. Evid Based Complement Alternat Med. 2011;2011:1–10.

31. Yang YS, Danis RP, Peterson RG, Dolan PL, Wu YQ. Acarbose partially inhibits microvascular retinopathy in the Zucker Diabetic Fatty rat (ZDF/Gmi-fa). Journal of ocular pharmacology and therapeutics. 2000;16(5):471–9.

32. Oltman CL, Coppey LJ, Gellett JS, Davidson EP, Lund DD, Yorek M a. Progression of vascular and neural dysfunction in sciatic nerves of Zucker diabetic fatty and Zucker rats. American journal of physiology Endocrinology and metabolism. 2005;289(1):E113–22.

33. Shimoshige Y, Ikuma K, Yamamoto T, Takakura S, Kawamura I, Seki J, et al. The effects of zenarestat, an aldose reductase inhibitor, on peripheral neuropathy in Zucker diabetic fatty rats. Metabolism: Clinical and Experimental. 2000;49(11):1395–9.

34. Wang B, Chandrasekera P, Pippin J. Leptin- and Leptin Receptor-Deficient Rodent Models: Relevance for Human Type 2 Diabetes. Current Diabetes Reviews. 2014;10(2):131–45.

35. Doria A, Patti M-E, Kahn CR. The emerging genetic architecture of type 2 diabetes. Cell metabolism. 2008;8(3):186–200.

36. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53.

37. Groop L, Pociot F. Genetics of diabetes - Are we missing the genes or the disease? Molecular and Cellular Endocrinology. 2014;382(1):726–39.

38. Watanabe RM. The genetics of insulin resistance: Where's Waldo? Current Diabetes Reports. 2010;10(6):476–84.

39. Das SK, Elbein SC. The Genetic Basis of Type 2 Diabetes. Cellscience. 2006;2(4):100–31.

40. Rich SS. Mapping genes in diabetes: Genetic epidemiological perspective. Diabetes. 1990;39(11):1315–9.

41. Kahn CR, Vicent D, Doria A. Non-Insulin-Dependent (Type-II) Diabetes Mellitus. Annual Review of Medicine. 1996;47:509–31.

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

42. Poulsen P, Ohm Kyvik K, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance - A population-based twin study. Diabetologia. 1999;42(2):139–45.

43. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature Genetics. 2008;40(5):638–45.

44. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007;316(5829):1336–41.

45. Saxena R, Voight BF, Lyssenko V, Burtt NP, Bakker PIW de, Chen H, et al. Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. Science. 2007;316(5829):1331–6.

46. PubMed. TCF7L2 transcription factor 7 like 2 [Rattus norvegicus (Norway rat)] [Internet]. [cited 2016 Nov 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/679869

47. PubMed. Pparg peroxisome proliferator-activated receptor gamma [Rattus norvegicus (Norway rat)] [Internet]. [cited 2016 Nov 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/25664

48. PubMed. Kcnj11 potassium voltage-gated channel subfamily J member 11 [Rattus norvegicus (Norway rat)] [Internet]. [cited 2016 Nov 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/83535

49. PubMed. TCF7L2 transcription factor 7 like 2 [Homo sapiens (human)] [Internet]. [cited 2016 Nov 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/6934

50. Prunier C. HBAHPH. Wnt signaling: Physiology and pathology. Growth Factors. 2004;22(3):141–50.

51. Yi F, Brubaker PL, Jin T. TCF-4 mediates cell type-specific regulation of proglucagon gene expression by β-catenin and glycogen synthase kinase-3β*. Journal of Biological Chemistry. 2005;280(2):1457–64.

52. Boj SF, Van Es JH, Huch M, Li VSW, José A, Hatzis P, et al. Diabetes risk gene and wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. Cell. 2012;151(7):1595–607.

53. Tong Y, Lin Y, Zhang YY, Yang J, Zhang YY, Liu H, et al. Association between TCF7L2gene polymorphisms and susceptibility to Type 2 Diabetes Mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. BMC Medical Genetics. 2009;10(1):15.

54. Grant SF a, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nature genetics. 2006;38(3):320–3.

55. Fehmann H-C, Göke R, Göke B. Cell and Molecular Biology of the Incretin Hormones Glucagon-Like Peptide-I and Glucose-Dependent Insulin Releasing Polypeptide. Endocrine Reviews. 1995;16(3):390.

56. Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. JClinInvest. 2007;117(8):2155–63.

57. Cauchi S, Meyre D, Dina C, Choquet H, Samson C, Gallina S, et al. Transcription factor TCF7L2 genetic study in the French population: expression in human beta-cells and adipose tissue and strong association with type 2 diabetes. Diabetes. 2006;55(October):2903–8.

58. Grant SFA. Understanding the elusive mechanism of action of TCF7L2 in metabolism. Diabetes. 2012;61(11):2657–8.

59. Kaminska D, Kuulasmaa T, Venesmaa S, Käkelä P, Vaittinen M, Pulkkinen L, et al. Adipose tissue TCF7L2 Splicing is regulated by weight loss and associates with glucose and fatty acid metabolism. Diabetes. 2012;61(11):2807–13.

60. McCarthy MI, Rorsman P, Gloyn AL. TCF7L2 and diabetes: A tale of two tissues, and of two species. Cell Metabolism. 2013;17(2):157–9.

61. PubMed. KCNJ11 potassium voltage-gated channel subfamily J member 11 [Homo sapiens (human)] [Internet]. [cited 2016 Nov 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/3767

62. Hani EH, Boutin P, Durand E, Inoue H, Permutt MA, Velho G, et al. Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): A meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. Diabetologia. 1998;41(12):1511–5.

63. Koster JC, Marshall BA, Ensor N, Corbett JA, Nichols CG. Targeted overactivity of beta cell K(ATP) channels induces profound neonatal diabetes. Cell. 2000;100(6):645–54.

64. PubMed. PPARG peroxisome proliferator activated receptor gamma [Homo sapiens (human)] [Internet]. [cited 2016 Nov 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/5468

65. Tontonoz P, Hu E, Spiegelman BM. Stimulation of adipogenesis in fibroblasts by PPARγ2, a lipid-activated transcription factor. Cell. 1994;79(7):1147–56.

66. Tontonoz P, Hu E, Graves RA, Budavari AI, Spiegelman BM. mPPARγ2: tissue specific regulator of an adipocyte enhancer. Genes & Development. 1994;8(10):1224–34.

67. Tontonoz P, Hu E, Spiegelman BM. Regulation of adipocyte gene expression and differentiation by peroxisome proliferator activated receptor γ. Current Opinion in Genetics & Development. 1995;5(5):571–6.

68. Werman A, Hollenberg A, Solanes G, Bjorbaek C, Vidal-Puig AJ, Flier JS. Ligand-independent Activation Domain in the N Terminus of Peroxisome Proliferator-activated Receptor (PPARγ): Differential Activity Of PPPARγ1 And -2 Isoforms And Influence Of Insulin. Journal of Biological Chemistry. 1997;272(32):20230–5.

69. Fajas L, Auboeuf D, Raspé E, Schoonjans K, Lefebvre AM, Saladin R, et al. The organization, promoter analysis, and expression of the human PPARgamma gene. The Journal of biological chemistry. 1997;272(30):18779–89.

70. Florez JC, Jablonski KA, Bayley N, Pollin TI, Bakker PIW de, Shuldiner AR, et al. TCF7L2 Polymorphisms and Progression to Diabetes in the Diabetes Prevention Program. N Engl J Med. 2006;355(3):241–50.

71. Schwanstecher C, Meyer U, Schwanstecher M. KIR6.2 Polymorphism Predisposes to T2D by Inducing Overactivity of β-Cell ATP-Sensitive K+ Channels. Diabetes. 2002;51(3):875–9.

72. Nielsen ED, Hansen L, Carstensen B, Echwald SM, Drivsholm T, Glümer C, et al. The E23K Varian of Kir6.2 Associates With Impaired Post-OGTT Serum Insulin Response and Increased Risk of Type 2 Diabetes. Diabetes. 2003;52(2):573–7.

73. Florez JC, Sjögren M, Burtt N, Orho-Melander M, Schayer S, Sun M, et al. Association Testing in 9,000 People Fails to Confirm the Association of the Insulin Receptor Substrate-1 G972R Polymorphism With Type 2 Diabetes. Diabetes. 2004;53(12):3313–8.

74. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007;316(June):1341–5.

75. Parton LE, Ye CP, Coppari R, Enriori PJ, Choi B, Zhang C-Y, et al. Glucose sensing by POMC neurons regulates glucose homeostasis and is impaired in obesity. Nature. 2007;449(7159):228–32.

76. Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, et al. A Pro12Ala substitution in PPAR gamma 2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nature Genetics. 1998;20(3):284–7.

77. Beamer BA, Yen CJ, Andersen RE, Muller D, Elahi D, Cheskin LJ, et al. Association of the Pro12Ala variant in the peroxisome proliferator- activated receptor-??2 gene with obesity in two Caucasian populations. Diabetes. 1998;47(11):1806–8.

78. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nature genetics. 2000;26(september):76–80.

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

79. Ruiz-Narvaez EA, Kraft P, Campos H. Ala 2 variant of the peroxisome proliferator-activated receptor-gamma gene (PPARG) is associated with higher polyunsaturated fat in adipose tissue and attenuates the protective effect of polyunsaturated fat intake on the risk of myocardial infarction. American Journal of Clinical Nutrition. 2007;86(4):1238–42.

80. Li Y, Zhu J, Ding JQ. Association of the PPARγ2 Pro12Ala polymorphism with increased risk of cardiovascular diseases. Genetics and Molecular Research. 2015;14(4):18662–74.

81. Taneera J, Lang S, Sharma A, Fadista J, Zhou Y, Ahlqvist E, et al. A Systems Genetics Approach Identifies Genes and Pathways for Type 2 Diabetes in Human Islets. Cell Metabolism. 2012;16(1):122–34.

82. Shu L, Sauter NS, Schulthess FT, Matveyenko A V. Transcription Factor 7-Like 2 Regulates B-cell Survival and Function in Human Pancreatic Islets. Diabetes. 2008;57(March):645–53.

83. Parton LE, McMillen PJ, Shen Y, Docherty E, Sharpe E, Diraison F, et al. Limited role for SREBP-1c in defective glucose-induced insulin secretion from Zucker diabetic fatty rat islets: a functional and gene profiling analysis. American journal of physiology Endocrinology and metabolism. 2006;291(5):E982-94.

84. Tokuyama Y, Fan Z, Furuta H, Makielski JC, Polonsky KS, Bell GI, et al. Rat inwardly rectifying potassium channel Kir6.2: cloning electrophysiological characterization, and decreased expression in pancreatic islets of male Zucker diabetic fatty rats. Biochemical and biophysical research communications. 1996;220(3):532–8.

85. Gyte A, Pritchard LE, Jones HB, Brennand JC, White A. Reduced expression of the KATP Channel Subunit, Kir6.2, is associated with decreased expression of neuropeptide Y and agouti-related protein in the hypothalami of zucker diabetic fatty rats. Journal of Neuroendocrinology. 2007;19(12):941–51.

86. Rosengren AH, Braun M, Mahdi T, Andersson SA, Travers ME, Shigeto M, et al. Reduced insulin exocytosis in human pancreatic β-cells with gene variants linked to type 2 diabetes. Diabetes. 2012;61(7):1726–33.

87. Parikh H, Lyssenko V, Groop LC. Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. BMC Medical Genomics. 2009;2(1):72.

88. Stratton IM, Adler AI, Neil HAW, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. British Medical Journal. 2000;321(7258):405–12.

89. UKPDS. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). The Lancet. 1998;352(9131):837–53.

90. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, et al. Management of hyperglycaemia in type 2 diabetes, 2015: a patient-centred approach. Update to a Position Statement of the American Diabetes Association and the European Association for the Study of Diabetes. Diabetologia. 2015;58(3):429–42.

91. American Diabetes Association. Standards of Medical care in diabetes - 2017. The Journal of Clinical and Applied Research and Education. 2017;40(January):1–142.

92. Ford ES, Zhao G, Li C. Pre-Diabetes and the Risk for Cardiovascular Disease. A Systematic Review of the Evidence. Journal of the American College of Cardiology. 2010;55(13):1310–7.

93. Ley SH, Korat AVA, Sun Q, Tobias DK, Zhang C, Qi L, et al. Contribution of the nurses' health studies to uncovering risk factors for type 2 diabetes: diet, lifestyle, biomarkers, and genetics. American Journal of Public Health. 2016;106(9):1624–30.

94. Frank B. Hu, Joann E Manson, Meir J Stampfer, Graham Colditz, Simin Liu CGS. Diet, lifestyle, and the risk of type 2 diabetes melitus in women. The New England Journal of Medicine. 2001;345(11):790–7.

95. Phillips MS, Liu Q, Hammond HA, Dugan V, Hey PJ, Caskey CJ, et al. Leptin receptor missense mutation in the fatty Zucker rat. Nature Genetics. 1996;13(1):18–9.

96. Griffen SC, Wang J, German MS. A genetic defect in beta-cell gene expression segregates independently from the fa locus in the ZDF rat.[In Process Citation]. Diabetes. 2001;50(1):63–8.

97. Sone M, Osamura RY. Leptin and the pituitary. Pituitary. 2001;4(1–2):15–23.

98. Paracchini V, Pedotti P, Taioli E. Genetics of leptin and obesity: A HuGE review. American Journal of Epidemiology. 2005;162(2):101–14.

99. Murugesan D, Arunachalam T, Ramamurthy V, Subramanian S. Association of polymorphisms in leptin receptor gene with obesity and type 2 diabetes in the local population of Coimbatore. Indian journal of human genetics. 2010;16(2):72–7.

100. Wang Z, Zhuo Q, Fu P, Piao J, Tian Y, Xu J, et al. Are the associations of plasma leptin and adiponectin with type 2 diabetes independent of obesity in older Chinese adults? Diabetes/Metabolism Research and Reviews. 2010;26(2):109–14.

101. Sempoux C, Guiot Y, Dubois D, Moulin P, Rahier J. Human type 2 diabetes: morphological evidence for abnormal beta-cell function. Diabetes. 2001;50(Suppl 1):S172–7.

102. Bonner-Weir S, O'Brien TD. Islets in type 2 diabetes: In honor of Dr. Robert C. Turner. Diabetes. 2008;57(11):2899–904.

103. Tomita T. Islet amyloid polypeptide in pancreatic islets from type 2 diabetic subjects. Islets. 2012;4(3):223–32.

104. Yagihashi S. Advances in pathology of diabetes from pancreatic islets to neuropathy - a tribute to Paul Langerhans. Pathology International. 2015 Apr 1;65(4):157–69.

105. Finegood DT, Mcarthur MD, Kojwang D, Thomas MJ, Topp BG, Leonard T, et al. Beta-cell mass Dynamics in Zucker Diabetic Fatty Rats. Rosiglitazone Prevents the Rise in Net Cell Death. Diabetes. 2001;50(6):1021–9.

106. Kakimoto T, Kimata H, Iwasaki S, Fukunari A, Utsumi H. Automated recognition and quantification of pancreatic islets in Zucker diabetic fatty rats treated with exendin-4. Journal of Endocrinology. 2013;216(1):13–20.

107. Jones HB, Bigley AL, Pemberton J, Randall KJ. Quantitative histopathological assessment of retardation of islets of langerhans degeneration in rosiglitazone-dosed obese ZDF rats using combined insulin and collagens (I and III) immunohistochemistry with automated image analysis and statistical modelin. Toxicologic pathology. 2013;41(3):425–44.

108. Janssen SW, Hermus AR, Lange WP, Knijnenburg Q, van der Laak JA, Sweep CG, et al. Progressive histopathological changes in pancreatic islets of Zucker Diabetic Fatty rats. Experimental and Clinical Endocrinology & Diabetes. 2001;109(5):273–82.

109. Inoue K, Hiramatsu S, Hisatomi A, Umeda F, Nawata H. Hypersecretion of amylin from the perfused pancreas of genetically obese (fa/fa) rats and its alteration with aging. Metabolism. 1993 May;42(5):654–8.

110. Nistala R, Raja A, Pulakat L. mTORC1 inhibitors rapamycin and metformin affect cardiovascular markers differentially in ZDF rats. Canadian journal of physiology and pharmacology. 2017;95(3):281–7.

111. Tsumura Y, Tsushima Y, Tamura A, Hasebe M, Kanou M, Kato H, et al. TMG-123, a novel glucokinase activator, exerts durable effects on hyperglycemia without increasing triglyceride in diabetic animal models. PloS one. 2017;12(2):1–16.

112. Othman EM, Oli RG, Arias-Loza PA, Kreissl MC, Stopper H. Metformin Protects Kidney Cells From Insulin-Mediated Genotoxicity In Vitro and in Male Zucker Diabetic Fatty Rats. Endocrinology. 2016;157(2):548–59.

113. Dong Y, Chen Y-T, Yang Y-X, Shou D, Li C-Y. Urinary Metabolomic Profiling in Zucker Diabetic Fatty Rats with Type 2 Diabetes Mellitus Treated with Glimepiride, Metformin, and Their Combination. Molecules. 2016;21(12):1446.

114. Chen W-P, Wang Y-D, Ma Y, Zhang Z-Y, Hu L-Y, Lin J-L, et al. Danhong Huayu Koufuye combined with metformin attenuated diabetic retinopathy in Zucker diabetic fatty. International Journal of Opthalmology. 2015;8(6):1094–100.

115. Forest T, Holder D, Smith A, Cunningham C, Yao X, Dey M, et al. Characterization of the exocrine pancreas in the male Zucker diabetic fatty rat model of type 2 diabetes mellitus following 3 months of treatment with sitagliptin. Endocrinology. 2014;155(3):783–92.

116. Reimer RA, Grover GJ, Koetzner L, Gahler RJ, Lyon MR, Wood S. Combining sitagliptin/metformin with a functional fiber delays diabetes progression in zucker rats. Journal of Endocrinology. 2014;220(3):361–73.

117. Wessels B, Ciapaite J, Van Den Broek NMA, Nicolay K, Prompers JJ. Metformin impairs mitochondrial function in skeletal muscle of both lean and diabetic rats in a Dose-dependent manner. PLoS ONE. 2014;9(6):1–10.

118. Konda VR, Desai A, Darland G, Grayson N, Bland JS. KDT501, a derivative from hops, normalizes glucose metabolism and body weight in rodent models of diabetes. PLoS ONE. 2014;9(1):1–11.

119. Kim YN, Kim S, Kim IY, Shin JH, Cho S, Yi SS, et al. Transcriptomic Analysis of Insulin-Sensitive Tissues from Anti-Diabetic Drug Treated ZDF Rats, a T2DM Animal Model. PLoS ONE. 2013;8(7):1–9.

120. Ito R, Tsujihata Y, Matsuda-Nagasumi K, Mori I, Negoro N, Takeuchi K. TAK-875, a GPR40/FFAR1 agonist, in combination with metformin prevents progression of diabetes and β-cell dysfunction in Zucker diabetic fatty rats. British Journal of Pharmacology. 2013;170(3):568–80.

121. Schmid PM, Resch M, Schach C, Birner C, Riegger G a, Luchner A, et al. Antidiabetic treatment restores adiponectin serum levels and APPL1 expression, but does not improve adiponectin-induced vasodilation and endothelial dysfunction in Zucker diabetic fatty rats. Cardiovascular diabetology. 2013;12(1):46.

122. Han SJ, Choi SE, Kang Y, Jung JG, Yi SA, Kim HJ, et al. Effect of sitagliptin plus metformin on β-cell function, islet integrity and islet gene expression in Zucker diabetic fatty rats. Diabetes Research and Clinical Practice. 2011;92(2):213–22.

123. Takiyama Y, Harumi T, Watanabe J, Fujita Y, Honjo J, Shimizu N, et al. Tubular injury in a rat model of type 2 diabetes is prevented by metformin: A possible role of HIF-1?? expression and oxygen metabolism. Diabetes. 2011;60(3):981–92.

124. Hwang IK, Kim IY, Joo EJ, Shin JH, Choi JW, Won MH, et al. Metformin normalizes type 2 diabetes-induced decrease in cell proliferation and neuroblast differentiation in the rat dentate gyrus. Neurochemical Research. 2010;35(4):645–50.

125. Forcheron F, Basset A, Carmine P Del, Beylot M. Lipase maturation factor 1: Its expression in Zucker diabetic rats, and effects of metformin and fenofibrate. Diabetes & Metabolism. 2009;35(35):452–7.

126. Shoghi KI, Finck BN, Schechtman KB, Herrero TSP, Gropler RJ, Welch MJ. In vivo metabolic phenotyping of myocardial substrate metabolism in rodents: Differential efficacy of metformin and rosiglitazone monotherapy. Circulation: Cardiovascular Imaging. 2009;2(5):373–81.

127. Forcheron F, Abdallah P, Basset A, del Carmine P, Haffar G, Beylot M. Nonalcoholic hepatic steatosis in Zucker diabetic rats: spontaneous evolution and effects of metformin and fenofibrate. Obesity (Silver Spring, Md). 2009;17(7):1381–9.

128. Metais C, Forcheron F, Abdallah P, Basset A, Del Carmine P, Bricca G, et al. Adiponectin receptors: expression in Zucker diabetic rats and effects of fenofibrate and metformin. Metabolism: Clinical and Experimental. 2008;57(7):946–53.

129. Smith AC, Mullen KL, Junkin KA, Nickerson J, Chabowski A, Bonen A, et al. Metformin and exercise reduce muscle FAT/CD36 and lipid accumulation and blunt the progression of high-fat diet-induced hyperglycemia. American journal of physiology Endocrinology and metabolism. 2007;293(1):E172-81.

130. Sreenan S, Sturis J, Pugh W, Burant CF, Polonsky KS. Prevention of hyperglycemia in the Zucker diabetic fatty rat by treatment with metformin or troglitazone. The American journal of physiology. 1996;271:E742–7.

131. Forcheron F, Basset A, Abdallah P, Del Carmine P, Gadot N, Beylot M. Diabetic cardiomyopathy: effects of fenofibrate and metformin in an experimental model--the Zucker diabetic rat. Cardiovascular diabetology. 2009;8(16):1–13.

132. Shang Q, Liu MK, Saumoy M, Holst JJ, Salen G, Xu G. The combination of colesevelam with sitagliptin enhances glycemic control in diabetic ZDF rat model. AJP: Gastrointestinal and Liver Physiology. 2012;302(8):G815–23.

133. Oguma T, Nakayama K, Kuriyama C, Matsushita Y, Yoshida K, Hikida K, et al. Intestinal Sodium Glucose Cotransporter 1 Inhibition Enhances Glucagon-Like Peptide-1 Secretion in Normal and Diabetic Rodents. The Journal of pharmacology and experimental therapeutics. 2015;354(3):279–89.

134. Takai S, Sakonjo H, Jin D. Significance of vascular dipeptidyl peptidase-4 inhibition on vascular protection in zucker diabetic Fatty rats. J Pharmacol Sci. 2014;125(4):386–93.

135. Marques C, Mega C, Gonçalves A, Rodrigues-Santos P, Teixeira-Lemos E, Teixeira F, et al. Sitagliptin prevents inflammation and apoptotic cell death in the kidney of type 2 diabetic animals. Mediators of Inflammation. 2014;2014:1–15.

136. Mega C, Vala H, Rodrigues-Santos P, Oliveira J, Teixeira F, Fernandes R, et al. Sitagliptin prevents aggravation of endocrine and exocrine pancreatic damage in the Zucker Diabetic Fatty rat - focus on amelioration of metabolic profile and tissue cytoprotective properties. Diabetology & metabolic syndrome. 2014;6:42.

137. Mega C, Teixeira De Lemos E, Vala H, Fernandes R, Oliveira J, Mascarenhas-Melo F, et al. Diabetic nephropathy amelioration by a low-dose sitagliptin in an animal model of type 2 diabetes (Zucker diabetic fatty rat). Experimental Diabetes Research. 2011;2011:1–12.

138. Gonçalves A, Leal E, Paiva A, Teixeira Lemos E, Teixeira F, Ribeiro CF, et al. Protective effects of the dipeptidyl peptidase IV inhibitor sitagliptin in the blood-retinal barrier in a type 2 diabetes animal model. Diabetes, Obesity and Metabolism. 2012;14(5):454–63.

139. Zhou W, Liu Z, Yao J, Chi F, Dong K, Yue X, et al. The Effects of Exenatide Microsphere on Serum BGP and ALP Levels in ZDF Rats after Implantation. Clinical Implant Dentistry and Related Research. 2013;(Dm):765–70.

140. Vrang N, Jelsing J, Simonsen L, Jensen AE, Thorup I, Søeborg H, et al. The effects of 13 wk of liraglutide treatment on endocrine and exocrine pancreas in male and female ZDF rats: a quantitative and qualitative analysis revealing no evidence of drug-induced pancreatitis. Am J Physiol Endocrinol Metab. 2012;303(2):E253-64.

141. Kwak HH, Shim WS, Hwang S, Son MK, Kim YJ, Kim TH, et al. Pharmacokinetics and efficacy of a biweekly dosage formulation of exenatide in zucker diabetic fatty (ZDF) rats. Pharmaceutical Research. 2009;26(11):2504–12.

142. Kwak HH, Shim WS, Son MK, Kim YJ, Kim TH, Youn HJ, et al. Efficacy of a new sustained-release microsphere formulation of exenatide, DA-3091, in Zucker diabetic fatty (ZDF) rats. European Journal of Pharmaceutical Sciences. 2010;40(2):103–9.

143. Gedulin BR, Smith P, Prickett KS, Tryon M, Barnhill S, Reynolds J, et al. Dose-response for glycaemic and metabolic changes 28 days after single injection of long-acting release exenatide in diabetic fatty Zucker rats. Diabetologia. 2005;48(7):1380–5.

144. Castillo GM, Reichstetter S, Bolotin EM. Extending residence time and stability of peptides by Protected Graft Copolymer (PGC) excipient: GLP-1 example. Pharmaceutical Research. 2012;29(1):306–18.

145. Tatarkiewicz K, Belanger P, Gu G, Parkes D, Roy D. No evidence of drug-induced pancreatitis in rats treated with exenatide for 13 weeks. Diabetes, obesity & metabolism. 2013;15(5):417–26.

146. Liu L, Wan J, Lang H, Si M, Zhu J, Zhou Y, et al. Dihydromyricetin delays the onset of hyperglycemia and ameliorates insulin resistance without excessive weight gain in Zucker diabetic fatty rats. Molecular and Cellular Endocrinology. 2017;439:105–15.

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

147. Beaudoin MS, Snook LA, Arkell AM, Stefanson A, Wan Z, Simpson JA, et al. Novel effects of rosiglitazone on SMAD2 and SMAD3 signaling in white adipose tissue of diabetic rats. Obesity. 2014;22(7):1632–42.

148. Nemanich S, Rani S, Shoghi K. In vivo multi-tissue efficacy of peroxisome proliferator-activated receptor-?? therapy on glucose and fatty acid metabolism in obese type 2 diabetic rats. Obesity. 2013;21(12):2522–9.

149. Schäfer HL, Linz W, Falk E, Glien M, Glombik H, Korn M, et al. AVE8134, a novel potent PPARα agonist, improves lipid profile and glucose metabolism in dyslipidemic mice and type 2 diabetic rats. Acta Pharmacologica Sinica. 2012;33(1):82–90.

150. Rohrbach K, Thomas MA, Glick S, Fung EN, Wang V, Watson L, et al. Ibipinabant attenuates β-cell loss in male Zucker diabetic fatty rats independently of its effects on body weight. Diabetes, Obesity and Metabolism. 2012;14(6):555–64.

151. Lu X, Guo X, Karathanasis SK, Zimmerman KM, Onyia JE, Peterson RG, et al. Rosiglitazone reverses endothelial dysfunction but not remodeling of femoral artery in Zucker diabetic fatty rats. Cardiovascular diabetology. 2010;9:19.

152. Wang X, Liu X, Zhan Y, Lavallie ER, Diblasio-Smith L, Collins-Racie L, et al. Pharmacogenomic, physiological, and biochemical investigations on safety and efficacy biomarkers associated with the peroxisome proliferator-activated receptor-gamma activator rosiglitazone in rodents: a translational medicine investigation. The Journal of pharmacology and experimental therapeutics. 2010;334(3):820–9.

153. Cosson E, Cohen-Boulakia F, Tarhzaoui K, Dabiré H, Léger G, Charnaux N, et al. Capillary endothelial but not lymphatic function is restored under rosiglitazone in Zucker Diabetic Fatty rats. Microvascular Research. 2009;77(2):220–5.

154. Mansour M, Coleman E, Dennis J, Akingbemi B, Schwartz D, Braden T, et al. Activation of PPAR by rosiglitazone does not negatively impact male sex steroid mormones in diabetic rats. PPAR Research. 2009;2009:1–8.

155. Iqbal MJ, Higginbotham A, Chickris N, Bollaert M, Rockway S, Banz WJ. A combination of CLA-DAG oil modifies the diabetic phenotype in male zucker diabetic fatty rats. Hormone and Metabolic Research. 2008;40(4):262–8.

156. Oltman CL, Davidson EP, Coppey LJ, Kleinschmidt TL, Lund DD, Adebara ET, et al. Vascular and neural dysfunction in Zucker diabetic fatty rats: A difficult condition to reverse. Diabetes, Obesity and Metabolism. 2008;10(1):64–74.

157. Banz WJ, Iqbal MJ, Bollaert M, Chickris N, James B, Higginbotham DA, et al. Ginseng modifies the diabetic phenotype and genes associated with diabetes in the male ZDF rat. Phytomedicine. 2007;14(10):681–9.

158. Gonzalez IC, Lamar J, Iradier F, Xu Y, Winneroski LL, York J, et al. Design and synthesis of a novel class of dual PPARgamma/delta agonists. Bioorganic & medicinal chemistry letters. 2007;17(4):1052–5.

159. Nadeau KJ, Ehlers LB, Aguirre LE, Reusch JEB, Draznin B. Discordance between intramuscular triglyceride and insulin sensitivity in skeletal muscle of Zucker diabetic rats after treatment with fenofibrate and rosiglitazone. Diabetes, Obesity and Metabolism. 2007;9(5):714–23.

160. Shi Q, Canada EJ, Xu Y, Warshawsky AM, Etgen GJ, Broderick CL, et al. Design and synthesis of novel and potent amide linked PPARgamma/delta dual agonists. Bioorganic & medicinal chemistry letters. 2007;17(24):6744–9.

161. Banz WWJ, Davis J, Steinle JJ, Adler S, Oitker J, Winters TA, et al. (+)-Z-Bisdehydrodoisynolic Acid Ameliorates Obesity and the Metabolic Syndrome in Female ZDF Rats. Obesity. 2005;13(11):1915–24.

162. Johns DG, Ao Z, Eybye M, Olzinski A, Costell M, Gruver S, et al. Rosiglitazone Protects against Ischemia / Reperfusion-Induced Leukocyte Adhesion in the Zucker Diabetic Fatty Rat. Pharmacology. 2005;315(3):1020–7.

163. Reifel-Miller A, Otto K, Hawkins E, Barr R, Bensch WR, Bull C, et al. A peroxisome proliferator-activated receptor alpha/gamma dual agonist with a unique in vitro profile and potent glucose and lipid effects in rodent models of type 2 diabetes and dyslipidemia. Molecular endocrinology. 2005;19(6):1593–605.

164. Winter CL, Lange JS, Davis MG, Gerwe GS, Downs TR, Peters KG, et al. A Nonspecific Phosphotyrosine Phosphatase Inhibitor , Bis (maltolato) oxovanadium (IV), Improves Glucose Tolerance and Prevents Diabetes in Zucker Diabetic Fatty Rats. Experimental Biology and Medicine. 2005;230(3):207–16.

165. Yue T, Bao W, Gu J, Cui J, Tao L, Ma X, et al. Rosiglitazone treatment in Zucker diabetic Fatty rats is associated with ameliorated cardiac insulin resistance and protection from ischemia/reperfusion-induced myocardial injury. Diabetes. 2005;54(2):554–62.

166. Banz WJ, Davis J, Peterson R, Iqbal MJ. Gene expression and adiposity are modified by soy protein in male Zucker diabetic fatty rats. Obesity research. 2004;12(12):1907–13.

167. Kuhlmann J, Neumann-Haefelin C, Belz U, Kalisch J, Juretschke HP, Stein M, et al. Intramyocellular lipid and insulin resistance: A longitudinal in vivo 1H-spectroscopic study in Zucker Diabetic Fatty rats. Diabetes. 2003;52(1):138–44.

168. Yuen VG, Bhanot S, Battell ML, Orvig C, McNeill JH. Chronic glucose-lowering effects of rosiglitazone and bis(ethylmaltolato)oxovanadium(IV) in ZDF rats. Canadian journal of physiology and pharmacology. 2003;81(11):1049–55.

169. Kim YY, Kang KM, Chung SH. Long-term administration of Sopungsungi-won (SP) prevents diabetic nephropathy in Zucker diabetic fatty rats. Archives of pharmacal research. 2002;25(6):917–22.

170. Kim YY, Kang HJ, Ko SK, Chung SH. Sopungsungi-won (SP) prevents the onset of hyperglycemia and hyperlipidemia in Zucker diabetic fatty rats. Arch Pharm Res. 2002;25(6):923–31.

171. Singh Ahuja H, Liu SHA, Crombie DL, Boehm M, Leibowitz MD, Heyman R a, et al. Differential effects of rexinoids and thiazolidinediones on metabolic gene expression in diabetic rodents. Molecular pharmacology. 2001;59(4):765–73.

172. Dana SL, Hoener PA, Bilakovics JM, Crombie DL, Ogilvie KM, Kauffman RF, et al. Peroxisome proliferator-activated receptor subtype-specific regulation of hepatic and peripheral gene expression in the Zucker diabetic fatty rat. Metabolism: Clinical and Experimental. 2001;50(8):963–71.

173. Way JM, Görgün CZ, Tong Q, Uysal KT, Brown KK, Harrington WW, et al. Adipose Tissue Resistin Expression is Severely Suppressed in Obesity and Stimulated by Peroxisome Proliferator-activated Receptor ?? Agonists. Journal of Biological Chemistry. 2001;276(28):25651–3.

174. Cai XJ, Lister CA, Buckingham RE, Pickavance L, Wilding J, Arch JRS, et al. Down-regulation of orexin gene expression by severe obesity in the rats: Studies in Zucker fatty and Zucker diabetic fatty rats and effects of rosiglitazone. Molecular Brain Research. 2000;77(1):131–7.

175. Maier VH, Melvin DR, Lister CA, Chapman H, Gould GW, Murphy GJ. v- and t-SNARE protein expression in models of insulin resistance: Normalization of glycemia by rosiglitazone treatment corrects overexpression of cellubrevin, vesicle-associated membrane protein-2, and syntaxin 4 in skeletal muscle of Zucker diabetic fat. Diabetes. 2000;49(4):618–25.

176. Smith SA, Lister CA, Toseland CDN, Buckingham RE. Rosiglitazone prevents the onset of hyperglycaemia and proteinuria in the Zucker diabetic fatty rat. Diabetes, Obesity and Metabolism. 2000;2(6):363–72.

177. Piercy V, Banner SE, Bhattacharyya A, Parsons AA, Sanger GJ, Smith SA, et al. Thermal, but not mechanical, nociceptive behavior is altered in the Zucker diabetic fatty rat and is independent of glycemic status. Journal of Diabetes and its Complications. 1999;13(3):163–9.

178. Brand CL, Sturis J, Gotfredsen CF, Fleckner J, Fledelius C, Hansen BF, et al. Dual PPARα/γ activation provides enhanced improvement of insulin sensitivity and glycemic control in ZDF rats. American Journal of Physiology - Endocrinology And Metabolism. 2003 Apr 1;284(4):E841 LP-E854.

179. Chen X, Osborne MC, Rybczynski PJ, Zeck R, Yang M, Xu J, et al. Pharmacological profile of a novel, non-TZD PPARg agonist. Diabetes, Obesity and Metabolism. 2005 Sep;7(5):536–46.

180. Kanda S, Nakashima R, Takahashi K, Tanaka J, Ogawa J, Ogata T, et al. Potent antidiabetic effects of rivoglitazone, a novel peroxisome proliferator-activated receptor-gamma agonist, in obese diabetic rodent models. Journal of pharmacological sciences. 2009;111(2):155–66.

181. Oguma T, Kuriyama C, Nakayama K, Matsushita Y, Yoshida K, Kiuchi S, et al. The effect of combined treatment with canagliflozin and teneligliptin on glucose intolerance in Zucker diabetic fatty rats. Journal of Pharmacological Sciences. 2015;127(4):456–61.

182. Watanabe Y, Nakayama K, Taniuchi N, Horai Y, Kuriyama C, Ueta K, et al. Beneficial effects of canagliflozin in combination with pioglitazone on insulin sensitivity in rodent models of obese type 2 diabetes. PLoS ONE. 2015;10(1):1–15.

183. Kuriyama C, Xu JZ, Lee SP, Qi J, Kimata H, Kakimoto T, et al. Analysis of the effect of canagliflozin on renal glucose reabsorption and progression of hyperglycemia in zucker diabetic Fatty rats. The Journal of pharmacology and experimental therapeutics. 2014;351(2):423–31.

184. Liang Y, Arakawa K, Ueta K, Matsushita Y, Kuriyama C, Martin T, et al. Effect of canagliflozin on renal threshold for glucose, glycemia, and body weight in normal and diabetic animal models. PLoS ONE. 2012;7(2):2–8.

185. Ueta K, O'Brien TP, McCoy G a, Kim K, Healey EC, Farmer TD, et al. Glucotoxicity targets hepatic glucokinase in Zucker diabetic fatty rats, a model of type 2 diabetes associated with obesity. American journal of physiology Endocrinology and metabolism. 2014;306(11):E1225-38.

186. Ferreira L, Teixeira-De-Lemos E, Pinto F, Parada B, Mega C, Vala H, et al. Effects of sitagliptin treatment on dysmetabolism, inflammation, and oxidative stress in an animal model of type 2 diabetes (ZDF rat). Mediators of Inflammation. 2010;2010:1–11.

187. Hansen HH, Jelsing J, Hansen CF, Hansen G, Vrang N, Mark M, et al. The sodium glucose cotransporter type 2 inhibitor empagliflozin preserves β-cell mass and restores glucose homeostasis in the male zucker diabetic fatty rat. The Journal of pharmacology and experimental therapeutics. 2014;350(3):657–64.

188. Yuen VG, Coleman J, Withers SG, Andersen RJ, Brayer GD, Mustafa S, et al. Glucose lowering effect of montbretin A in Zucker Diabetic Fatty rats. Molecular and Cellular Biochemistry. 2016;411(1–2):373–81.

189. Sturis J, Pugh WL, Tang J, Polonsky KS. Prevention of diabetes does not completely prevent insulin secretory defects in the ZDF rat. Am J Physiol. 1995;269(4 Pt 1):E786-92.

190. Friedman JE, de Venté JE, Peterson RG, Dohm GL. Altered expression of muscle glucose transporter GLUT-4 in diabetic fatty Zucker rats (ZDF/Drt-fa). The American journal of physiology. 1991;261(6 Pt 1):E782-8.

191. Dhalla a K, Santikul M, Chisholm JW, Belardinelli L, Reaven GM. Comparison of the antilipolytic effects of an A1 adenosine receptor partial agonist in normal and diabetic rats. Diabetes, obesity & metabolism. 2009;11(2):95–101.

192. Dang Q, Kasibhatla SR, Reddy KR, Jiang T, Reddy MR, Potter SC, et al. Discovery of Potent and Specific Fructose-1,6-Bisphosphatase Inhibitors and a Series of Orally Bioavailable Phosphoramidase Sensitive Prodrugs for the Treatment of Type 2 Diabetes. Journal of the American Chemical Society. 2007;.(1):1047–53.

193. Erion MD, van Poelje PD, Dang Q, Kasibhatla SR, Potter SC, Reddy MR, et al. MB06322 (CS-917): A potent and selective inhibitor of fructose 1,6-bisphosphatase for controlling gluconeogenesis in type 2 diabetes. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(22):7970–5.

194. Van Poelje PD, Potter SC, Chandramouli VC, Landau BR, Dang Q, Erion MD. Inhibition of fructose 1,6-bisphosphatase reduces excessive endogenous glucose production and attenuates hyperglycemia in Zucker diabetic fatty rats. Diabetes. 2006;55(6):1747–54.

195. Hummel KP, Dickie MM, Coleman DL. Diabetes, a New Mutation in the Mouse. Science. 1966;153(3740):1127–8.

196. The Jackson Laboratory. Mouse Strain Datasheet 000642 [Internet]. [cited 2017 Jun 22]. Available from: https://www.jax.org/strain/000642

197. Hummel KP, Coleman DL, Lane PW. The influence of genetic background on expression of mutations at the diabetes locus in the mouse. I. C57BL/KsJ and C57BL/6J strains. Biochemical Genetics. 1972;7(1):1–13.

198. Boquist L, Hellman B, Lernmark A, Täljedal IB. Influence of the mutation "diabetes" on insulin release and islet morphology in mice of different genetic backgrounds. Journal of Cell Biology. 1974;62(1):77–89.

199. Coleman DL, Jackson T. Hyperinsulinemia in Pre-Weaning Diabetes (db) Mice. Diabetologia. 1974;10(Suppl 1):607–10.

200. Coleman DL. Diabetes-obesity syndromes in mice. Diabetes. 1982;31(Suppl. 1):1–6.

201. Coleman DL, Hummel KP. Studies with the mutation, diabetes, in the mouse. Diabetologia. 1967;3(2):238–48.

202. Berglund O, Frankel BJ, Hellman B. Development of the insulin secretory defect in genetically diabetic (db/db) mouse. Acta endocrinologica. 1978;87(3):543–51.

203. Like AA, Chick WL. Studies in the diabetic mutant mouse: I. Light microscopy and radioautography of pancreatic islets. Diabetologia. 1970;6(3):207–15.

204. Like AA, Chick WL. Studies in the diabetic mutant mouse: II. Electron Microscopy of Pancreatic Islets. Diabetologia. 1970;6(3):216–42.

205. Chick WL, Like AA. Studies in the diabetic mutant mouse: III. Physiological factors associated with alterations in beta cell proliferation. Diabetologia. 1970;6(3):243–51.

206. Tuman RW, Doisy RJ. The Influence of Age on the Development of Hypertriglyceridaemia and Hypercholesterolaemia in Genetically Diabetic Mice. Diabetologia. 1977 Jan;13(1):7–11.

207. Coleman DL, Hummel KP. Effects of parabiosis of normal with genetically diabetic mice. The American journal of physiology. 1969;217(5):1298–304.

208. Wyse BM, Dulin WE. The influence of age and dietary conditions on diabetes in the db mouse. Diabetologia. 1970;6(3):268–73.

209. Malaisse WJ, Malaisse-Lagae F, Coleman DL. Insulin Secretion in Mice with an Hereditary Diabetes. Experimental Biology and Medicine. 1968;129(1):65–9.

210. Nishina PM, Lowe S, Wang J, Paigen B. Characterization of plasma lipids in genetically obese mice: The mutants obese, diabetes, fat, tubby, and lethal yellow. Metabolism. 1994;43(5):549–53.

211. Grounds MD, Radley HG, Lynch GS, Nagaraju K, De Luca A. Towards developing standard operating procedures for pre-clinical testing in the mdx mouse model of Duchenne muscular dystrophy. Neurobiology of Disease. 2008;31(1):1–19.

212. Dutta S, Sengupta P. Men and mice: Relating their ages. Life Sciences. 2016;152:244–8.

213. Like AA, Lavine RL, Poffenbarger PL, Chick WL. Studies in the Diabetic Mutant Mouse: VI. Evolution of Glomerular Lesions and Associated Proteinuria. The American Journal of Pathology. 1972;66(2):193–224.

214. Gaertner K. Glomerular hyperfiltration during the onset of diabetes mellitus in two strains of diabetic mice (c57bl/6j db/db and c57bl/ksj db/db). Diabetologia. 1978;15(1):59–63.

215. Meade CJ, Brandon DR, Smith W, Simmonds RG, Harris S, Sowter C. The relationship between hyperglycaemia and renal immune complex deposition in mice with inherited diabetes. Clin Exp Immunol. 1981;43(1):109–20.

216. Kitada M, Kume S, Imaizumi N, Koya D. Resveratrol improves oxidative stress and protects against diabetic nephropathy through normalization of Mn-SOD dysfunction in AMPK/SIRT1- independent pathway. Diabetes. 2011;60(2):634–43.

217. Susztak K, Raff AC, Schiffer M, Böttinger EP. Glucose-Induced Reactive Oxygen Species Cause Apoptosis of Podocytes and Podocyte Depletion at the Onset of Diabetic Nephropathy. Diabetes. 2006;55(1):225–33.

218. Koya D, Haneda M, Nakagawa H, Isshiki K, Sato H, Maeda S, et al. Amelioration of accelerated diabetic mesangial expansion by treatment with a PKC beta inhibitor in diabetic db/db mice, a rodent model for type 2 diabetes. FASEB. 2000;14(3):439–47.

219. Midena E, Segato T, Radin S, Di Giorgio G, Meneghini F, Piermarocchi S, et al. Studies on the retina of the diabetic db/db mouse: I. endothelial cell-pericyte ratio. Ophthalmic Research. 1989;21(2):106–11.

220. Clements RS, Robison WG, Cohen MP. Anti-glycated albumin therapy ameliorates early retinal microvascular pathology in db/db mice. Journal of Diabetes and its Complications. 1998;12(1):28–33.

221. Cheung AKH, Fung MKL, Lo ACY, Lam TTL, So KF, Chung SSM, et al. Aldose Reductase Deficiency Prevents Diabetes-Induced Blood-Retinal Barrier Breakdown, Apoptosis, and Glial Reactivation in the Retina of db/db Mice. Diabetes. 2005;54(11):3119–25.

222. Tadayoni R, Paques M, Gaudric A, Vicaut E. Erythrocyte and leukocyte dynamics in the retinal capillaries of diabetic mice. Experimental Eye Research. 2003;77(4):497–504.

223. Li J, Wang JJ, Yu Q, Chen K, Mahadev K, Zhang SX. Inhibition of Reactive Oxygen Species by Lovastatin Downregulates Vascular Endothelial Growth Factor Expression and Ameliorates Blood Retinal Barrier Breakdown in db/db Mice. Diabetes. 2010;59(June):1528–38.

224. Cohen MP, Hud E, Shea E, Shearman CW. Vitreous Fluid of db/db Mice Exhibits Alterations in Angiogenic and Metabolic Factors Consistent with Early Diabetic Retinopathy. Ophthalmic Research. 2008;40(1):5–9.

225. Sima AAF, Robertson DM. Peripheral neuropathy in mutant diabetic mouse [C57BL/Ks(db/db)]. Acta Neuropathologica. 1978;41(2):85–9.

226. Robertson DM, Sima AAF. Diabetic neuropathy in the mutant mouse [C57BL/ks9(db/db)] - A morphometric study. Diabetes. 1980;29(1):60–7.

227. Hanker JS, Ambrose WW, Yates PE, Koch G, Carson KA. Peripheral Neuropathy in Mouse Hereditary Diabetes Mellitus *. Acta Neuropathologica. 1980;51:145–53.

228. Sullivan KA, Hayes JM, Wiggin TD, Backus C, Su Oh S, Lentz SI, et al. Mouse models of diabetic neuropathy. Neurobiology of Disease. 2007;28(3):276–85.

229. Moore SA, Peterson RG, Felten DL, Cartwright TR, O'Connor BL. Reduced sensory and motor conduction velocity in 25-week-old diabetic [C57BL Ks (db db)] mice. Experimental Neurology. 1980;70(3):548–55.

230. PubMed. Tcf7l2 transcription factor 7 like 2, T cell specific, HMG box [Mus musculus (house mouse)] [Internet]. [cited 2017 Jul 5]. Available from: https://www.ncbi.nlm.nih.gov/gene/21416

231. PubMed. Kcnj11 potassium inwardly rectifying channel, subfamily J, member 11 [Mus musculus (house mouse)] [Internet]. [cited 2017 Jul 5]. Available from: https://www.ncbi.nlm.nih.gov/gene/16514

232. PubMed. Pparg peroxisome proliferator activated receptor gamma [Mus musculus (house mouse)] [Internet]. [cited 2017 Jul 5]. Available from: https://www.ncbi.nlm.nih.gov/gene/19016

233. Chen ZL, Shao WJ, Xu F, Liu L, Lin BS, Wei XH, et al. Acute Wnt pathway activation positively regulates leptin gene expression in mature adipocytes. Cellular Signalling. 2015;27(3):587–97.

234. Shu L, Matveyenko A V., Kerr-Conte J, Cho JH, McIntosh CHS, Maedler K. Decreased TCF7L2 protein levels in type 2 diabetes mellitus correlate with downregulation of GIP- and GLP-1 receptors and impaired beta-cell function. Human Molecular Genetics. 2009;18(13):2388–99.

235. Sharma M, Mohapatra J, Malik U, Nagar J, Chatterjee A, Ramachandran B, et al. Effect of pioglitazone on metabolic features in endotoxemia model in obese diabetic db/db mice. Journal of Diabetes. 2017;9(6):613–21.

236. Singh VP, Gurunathan C, Singh S, Singh B, Lakshmi BJ, Mishra AP, et al. Genetic deletion of Wdr13 improves the metabolic phenotype of Lepr db/db mice by modulating AP1 and PPARγ target genes. Diabetologia. 2015;58(2):384–92.

237. Su W, Xie Z, Guo Z, Duncan MJ, Lutshumba J, Gong MC. Altered clock gene expression and vascular smooth muscle diurnal contractile variations in type 2 diabetic db/db mice. American Journal of Physiology, Heart and Circulatory Physiology. 2012;302(3):H621-33.

238. Chen H, Charlat O, Tartaglia LA, Woolf EA, Weng X, Ellis SJ, et al. Evidence that the diabetes gene encodes the leptin receptor: Identification of a mutation in the leptin receptor gene in db/db mice. Cell. 1996;84(3):491–5.

239. Tartaglia LA, Dembski M, Weng X, Deng N, Culpepper J, Devos R, et al. Identification and expression cloning of a leptin receptor, OB-R. Cell. 1995;83(7):1263–71.

240. Coleman DL. Effects of parabiosis of obese with diabetes and normal mice. Diabetologia. 1973;9(4):294–8.

241. Chlouverakis C. Insulin resistance of parabiotic obese-hyperglycemic mice. Horm Metab Res. 1972;1972(4):143–8.

242. Haessler HA, Crawford JD. Alterations in the Fatty Acid Composition of Depot Fat Associated With Obesity. Annals of the New York Academy of Sciences. 1965;131(1):476–84.

243. Campfield LA, Smith FJ, Guisez Y, Devos R, Campfield LA, Smith FJ, et al. Recombinant Mouse OB Protein: Evidence for a Peripheral Signal Linking Adiposity and Central Neural Networks. Science. 1995;269(5223):546–9.

244. Halaas J, Gajiwala K, Maffei M, Cohen S, Chait B, Rabinowitz D, et al. Weight-reducing effects of the plasma protein encoded by the obese gene. Science. 1995;269(5223):543–6.

245. Pelleymounter MA, Cullen MJ, Baker MB, Hecht R, Pelleymounter MA, Cullen MJ, et al. Effects of the Obese Gene Product on Body Weight Regulation in Ob/Ob Mice. Science. 1995;269(5223):540–3.

246. Stephens TW, Basinski M, Bristow PK, Bue-Valleskey JM, Burgett SG, Craft L, et al. The role of neuropeptide Y in the antiobesity action of the obese gene product. Vol. 377, Nature. 1995. p. 530–2.

247. Takada K, Kanatsuka A, Tokuyama Y, Yagui K, Nishimura M, Saito Y, et al. Islet amyloid polypeptide/amylin contents in pancreas change with increasing age in genetically obese and diabetic mice. Diabetes Research and Clinical Practice. 1996;33(3):153–8.

248. Chen F, Dong RR, Zhong KL, Ghosh A, Tang SS, Long Y, et al. Antidiabetic drugs restore abnormal transport of amyloid-β across the blood-brain barrier and memory impairment in db/db mice. Neuropharmacology. 2016;101:123–36.

249. Li DJ, Huang F, Lu WJ, Jiang GJ, Deng YP, Shen FM. Metformin promotes irisin release from murine skeletal muscle independently of AMP-activated protein kinase activation. Acta Physiologica. 2015;213(3):711–21.

250. Tang T, Reed MJ. Exercise adds to metformin and acarbose efficacy in db/db mice. Metabolism: Clinical and Experimental. 2001;50(9):1049–53.

251. Cheng L, Song J, Li G, Liu Y, Wang Y, Meng X, et al. Effects of the Tangningtongluo formula as an alternative strategy for diabetics via upregulation of insulin receptor substrate-1. Molecular Medicine Reports. 2017;16(1):703–9.

252. Dludla P V., Muller CJF, Joubert E, Louw J, Essop MF, Gabuza KB, et al. Aspalathin protects the heart against hyperglycemia-induced oxidative damage by up-regulating Nrf2 expression. Molecules. 2017;22(1):1–16.

253. Huang MQ, Zhou CJ, Zhang YP, Zhang XQ, Xu W, Lin J, et al. Salvianolic Acid B Ameliorates Hyperglycemia and Dyslipidemia in db/db Mice through the AMPK Pathway. Cellular Physiology and Biochemistry. 2016;40(5):933–43.

254. Bae U-J, Choi E-K, Oh M-R, Jung S-J, Park J, Jung T-S, et al. Angelica gigas Ameliorates Hyperglycemia and Hepatic Steatosis in C57BL/KsJ-db/db Mice via Activation of AMP-Activated Protein Kinase Signaling Pathway. The American journal of Chinese medicine. 2016;44(8):1627–38.

255. Adam J, Brandmaier S, Leonhardt J, Scheerer MF, Mohney RP, Xu T, et al. Metformin effect on nontargeted metabolite profiles in patients with type 2 diabetes and in multiple murine tissues. Diabetes. 2016;65(12):3776–85.

256. Mitchell PL, Nachbar R, Lachance D, St-Pierre P, Trottier J, Barbier O, et al. Treatment with a novel agent combining docosahexaenoate and metformin increases protectin DX and IL-6 production in skeletal muscle and reduces insulin resistance in obese diabetic db/db mice. Diabetes, Obesity and Metabolism. 2017;19(3):313–9.

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

257. Gallo LA, Ward MS, Fotheringham AK, Zhuang A, Borg DJ, Flemming NB, et al. Erratum: Once daily administration of the SGLT2 inhibitor, empagliflozin, attenuates markers of renal fibrosis without improving albuminuria in diabetic db/db mice. Scientific reports. 2016;6(January):26428.

258. Bao Q, Shen X, Qian L, Gong C, Nie M, Dong Y. Anti-diabetic activities of catalpol in db/db mice. Korean Journal of Physiology and Pharmacology. 2016;20(2):153–60.

259. Asare GA, Adjei S, Afriyie D, Appiah-Danquawh AB, Asia J, Asiedu B, et al. Croton membranaceus improves some biomarkers of cardiovascular disease and diabetes in genetic animal models. Journal of Clinical and Diagnostic Research. 2015;9(12):OF01–5.

260. Solini A, Rossi C, Duranti E, Taddei S, Natali A, Virdis A. Saxagliptin prevents vascular remodeling and oxidative stress in db/db mice. Role of endothelial nitric oxide synthase uncoupling and cyclooxygenase. Vascular Pharmacology. 2016;76:62–71.

261. Kim D Il, Park MJ, Heo YR, Park SH. Metformin ameliorates lipotoxicity-induced mesangial cell apoptosis partly via upregulation of glucagon like peptide-1 receptor (GLP-1R). Archives of Biochemistry and Biophysics. 2015;584:90–7.

262. Wang X, Chen Y, Abdelkader D, Hassan W, Sun H, Liu J. Combination therapy with oleanolic acid and metformin as a synergistic treatment for diabetes. Journal of Diabetes Research. 2015;2015:1–12.

263. Cheng L, Meng XB, Lu S, Wang TT, Liu Y, Sun GB, et al. Evaluation of Hypoglycemic Efficacy of Tangningtongluo Formula, a Traditional Chinese Miao Medicine, in Two Rodent Animal Models. Journal of Diabetes Research. 2014;2014:1–12.

264. Somineni HK, Boivin GP, Elased KM. Daily exercise training protects against albuminuria and angiotensin converting enzyme 2 shedding in db/db diabetic mice. Journal of Endocrinology. 2014;221(2):235–51.

265. Kim MH, Jee JH, Park S, Lee MS, Kim KW, Lee MK. Metformin enhances glucagon-like peptide 1 via cooperation between insulin and Wnt signaling. Journal of Endocrinology. 2014;220(2):117–28.

266. Anthony J, Kelkar A, Wilankar C, Ranjith V, Bhumra SK, Mutt S, et al. Discovery of P1736, a Novel Antidiabetic Compound That Improves Peripheral Insulin Sensitivity in Mice Models. PLoS ONE. 2013;8(10):1–10.

267. Ly A, Scheerer MF, Zukunft S, Muschet C, Merl J, Adamski J, et al. Retinal proteome alterations in a mouse model of type 2 diabetes. Diabetologia. 2014;57(1):192–203.

268. Bruckbauer A, Zemel MB. Effects of metformin, resveratrol, and hydroxymethylbutyrate on insulin sensitivity. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy. 2013;6:93–102.

269. Tripp ML, Darland G, Konda VR, Pacioretty LM, Chang JL, Bland JS, et al. Optimized mixture of hops rho iso-alpha acids-rich extract and acacia proanthocyanidins- rich extract reduces insulin resistance in 3t3-l1 adipocytes and improves glucose and insulin control in db/db mice. Nutrition Research and Practice. 2012;6(5):405–13.

270. Li J, Deng J, Sheng W, Zuo Z. Metformin attenuates Alzheimer's disease-like neuropathology in obese, leptin-resistant mice. Pharmacology Biochemistry and Behavior. 2012;101(4):564–74.

271. Caton PW, Kieswich J, Yaqoob MM, Holness MJ, Sugden MC. Metformin opposes impaired AMPK and SIRT1 function and deleterious changes in core clock protein expression in white adipose tissue of genetically-obese db/db mice. Diabetes, obesity & metabolism. 2011;13(12):1097–104.

272. Akanksha, Srivastava AK, Maurya R. Antihyperglycemic activity of compounds isolated from indian medicinal plants. Indian Journal of Experimental Biology. 2010;48(3):294–8.

273. Caton PW, Nayuni NK, Kieswich J, Khan NQ, Yaqoob MM, Corder R. Metformin suppresses hepatic gluconeogenesis through induction of SIRT1 and GCN5. Journal of Endocrinology. 2010;205(1):97–106.

274. Maurya R, Akanksha, Jayendra, Singh AB, Srivastava AK. Coagulanolide, a withanolide from Withania coagulans fruits and antihyperglycemic activity. Bioorganic and Medicinal Chemistry Letters. 2008;18(24):6534–7.

275. Heishi M, Hayashi K, Ichihara J, Ishikawa H, Kawamura T, Kanaoka M, et al. Comparison of gene expression changes induced by biguanides in db/db mice liver. The Journal of toxicological sciences. 2008;33(3):339–47.

276. Yoon SH, Han EJ, Sung JH, Chung SH. Anti-diabetic effects of compound K versus metformin versus compound K-metformin combination therapy in diabetic db/db mice. Biological & pharmaceutical bulletin. 2007;30(11):2196–200.

277. Heishi M, Ichihara J, Teramoto R, Itakura Y, Hayashi K, Ishikawa H, et al. Global gene expression analysis in liver of obese diabetic db/db mice treated with metformin. Diabetologia. 2006;49(7):1647–55.

278. Fujita H, Fujishima H, Koshimura J, Hosoba M, Yoshioka N, Shimotomai T, et al. Effects of antidiabetic treatment with metformin and insulin on serum and adipose tissue adiponectin levels in db/db mice. Endocrine journal. 2005;52(4):427–33.

279. Fujita H, Fujishima H, Morii T, Koshimura J, Narita T, Kakei M, et al. Effect of metformin on adipose tissue resistin expression in db / db mice. Biochemical and biophysical research communications. 2002;298(3):345–9.

280. Ono M, Itakura Y, Nonomura T, Nakagawa T, Nakayama C, Taiji M, et al. Intermittent administration of brain-derived neurotrophic factor ameliorates glucose metabolism in obese diabetic mice. Metabolism. 2000;49(1):129–33.

281. Raza S, Srivastava SP, Srivastava DS, Srivastava AK, Haq W, Katti SB. Thiazolidin-4-one and thiazinan-4-one derivatives analogous to rosiglitazone as potential antihyperglycemic and antidyslipidemic agents. European Journal of Medicinal Chemistry. 2013;63:611–20.

282. Kushwaha RN, Srivastava R, Mishra A, Rawat AK, Srivastava AK, Haq W, et al. Design, Synthesis, Biological Screening, and Molecular Docking Studies of Piperazine-Derived Constrained Inhibitors of DPP-IV for the Treatment of Type 2 Diabetes. Chem Biol Drug Des. 2015;85(4):439–46.

283. Zhao SJ, Wang DH, Li YW, Han L, Xiao X, Ma M, et al. A novel selective VPACAC2 agonist peptide-conjugated chitosan modified selenium nanoparticles with enhanced anti-type 2 diabetes synergy effects. International Journal of Nanomedicine. 2017;12:2143–60.

284. Hernández C, Bogdanov P, Corraliza L, García-Ramírez M, Solà-Adell C, Arranz JA, et al. Topical administration of GLP-1 receptor agonists prevents retinal neurodegeneration in experimental diabetes. Diabetes. 2016;65(1):172–87.

285. Hou S, Li C, Huan Y, Liu S, Liu Q, Sun S, et al. Effects of E2HSA, a long-acting glucagon like peptide-1 receptor agonist, on glycemic control and beta cell function in spontaneous diabetic db/db mice. Journal of Diabetes Research. 2015;2015:1–17.

286. Nakashima R, Yano T, Ogawa J, Tanaka N, Toda N, Yoshida M, et al. Potentiation of insulin secretion and improvement of glucose intolerance by combining a novel G protein-coupled receptor 40 agonist DS-1558 with glucagon-like peptide-1 receptor agonists. European Journal of Pharmacology. 2014;737:194–201.

287. Liang R, Li X, Zhang R, Shi Y, Wang A, Chen D, et al. Acylation of exenatide by glycolic acid and its anti-diabetic activities in db/db mice. Pharmaceutical research. 2014;31(8):1958–66.

288. Zhang B, He D, Fan Y, Liu N, Chen Y. Oral delivery of exenatide via microspheres prepared by cross-linking of alginate and hyaluronate. PLoS ONE. 2014;9(1):1–7.

289. Yu L, Li K, Liu X, Chen C, Bao Y, Ci T, et al. In vitro and in vivo evaluation of a once-weekly formulation of an antidiabetic peptide drug exenatide in an injectable thermogel. Journal of Pharmaceutical Sciences. 2013;102(11):4140–9.

290. Dhanesha N, Joharapurkar A, Shah G, Kshirsagar S, Patel V, Patel K, et al. Treatment with exendin-4 improves the antidiabetic efficacy and reverses hepatic steatosis in glucokinase activator treated db/db mice. European Journal of Pharmacology. 2013;714(1–3):188–92.

291. Fosgerau K, Jessen L, Lind Tolborg J, Østerlund T, Schæffer Larsen K, Rolsted K, et al. The novel GLP-1-gastrin dual agonist, ZP3022, increases β-cell mass and prevents diabetes in db/db mice. Diabetes, Obesity and Metabolism. 2013;15(1):62–71.

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

292. Liu HY, Chung C, Yang W, Liang C, Wang C, Chang C, et al. Exendin-4 improves resistance to Listeria monocytogenes infection in diabetic db/db mice. Journal of Veterinary Science. 2012;13(3):245–52.

293. Rieg T, Gerasimova M, Murray F, Masuda T, Tang T, Rose M, et al. Natriuretic effect by exendin-4, but not the DPP-4 inhibitor alogliptin, is mediated via the GLP-1 receptor and preserved in obese type 2 diabetic mice. Am J Physiol Renal Physiol. 2012;303(7):F963-71.

294. Patel V, Joharapurkar A, Gandhi T, Patel K, Dhanesha N, Kshirsagar S, et al. Omeprazole improves anti-obesity and antidiabetic effects of exendin-4 in db/db mice. Journal of Diabetes. 2013;5(2):163–71.

295. Kan M, Guo G, Singh B, Singh V, Zochodne DW. Glucagon-Like Peptide 1, Insulin, Sensory Neurons, and Diabetic Neuropathy. Journal of Neuropathology & Experimental Neurology. 2012;71(6):494–510.

296. Dhanesha N, Joharapurkar A, Shah G, Dhote V, Kshirsagar S, Bahekar R, et al. Exendin-4 reduces glycemia by increasing liver glucokinase activity: An insulin independent effect. Pharmacological Reports. 2012;64(1):140–9.

297. Liu BY, Jiang Y, Liu Z, Li S, Debing LU, Chen B. Down-regulation of zinc transporter 8 in the pancreas of db/db mice is rescued by Exendin-4 administration. Molecular Medicine Reports. 2011;4(1):47–52.

298. Hamilton A, Patterson S, Porter D, Gault VA, Holscher C. Novel GLP-1 mimetics developed to treat type 2 diabetes promote progenitor cell proliferation in the brain. Journal of Neuroscience Research. 2011;89(4):481–9.

299. Gedulin BR, Smith PA, Jodka CM, Chen K, Bhavsar S, Nielsen LL, et al. Pharmacokinetics and pharmacodynamics of exenatide following alternate routes of administration. International Journal of Pharmaceutics. 2008;356(1–2):231–8.

300. Park CW, Kim HW, Ko SH, Lim JH, Ryu GR, Chung HW, et al. Long-Term Treatment of Glucagon-Like Peptide-1 Analog Exendin-4 Ameliorates Diabetic Nephropathy through Improving Metabolic Anomalies in db/db Mice. Journal of the American Society of Nephrology. 2007;18(4):1227–38.

301. Wang Q, Brubaker P. Glucagon-like peptide-1 treatment delays the onset of diabetes in 8 week-old db/db mice. Diabetologia. 2002;45(9):1263–73.

302. Rolin B, Larsen MO, Gotfredsen CF, Deacon CF, Carr RD, Wilken M, et al. The long-acting GLP-1 derivative NN2211 ameliorates glycemia and increases beta-cell mass in diabetic mice. American Journal of Physiology: Endocrinology and Metabolism. 2002;283(4):E745-52.

303. Yoshida S, Ohishi T, Matsui T, Tanaka H, Oshima H, Yonetoku Y, et al. The role of small molecule GPR119 agonist, AS1535907, in glucose-stimulated insulin secretion and pancreatic β-cell function. Diabetes, obesity & metabolism. 2011;13(1):34–41.

304. Xiong WT, Gu L, Wang C, Sun HX, Liu X. Anti-hyperglycemic and hypolipidemic effects of Cistanche tubulosa in type 2 diabetic db/db mice. Journal of Ethnopharmacology. 2013;150(3):935–45.

305. Combs TP, Wagner J JAB, Doebber T, Wang WJ, Zhang BB, Tanen M, et al. Induction of adipocyte complement related protein of 30 kilodaltons by PPAR-gamma agnoists: a potential mechanism of insulin sensitization. Endocrinology. 2002;143(September):998–1007.

306. Lejnev K, Khomsky L, Bokvist K, Mistriel-Zarbib S, Naveh T, Farb TB, et al. Thioredoxin-mimetic peptides (TXM) inhibit inflammatory pathways associated with high-glucose and oxidative stress. Free Radical Biology and Medicine. 2016;99:557–71.

307. Choi SI, Lee HA, Han JS. Gynura procumbens extract improves insulin sensitivity and suppresses hepatic gluconeogenesis in C57BL/KsJ-db/db mice. Nutrition Research and Practice. 2016;10(5):507–15.

308. Choi KH, Lee HA, Park MH, Han J-S. Mulberry (Morus alba L.) Fruit Extract Containing Anthocyanins Improves Glycemic Control and Insulin Sensitivity via Activation of AMP-Activated Protein Kinase in Diabetic C57BL/Ksj-db/db Mice. Journal of Medicinal Food. 2016;19(8):737–45.

309. Yuan X, Li H, Bai H, Zhao X, Zhang C, Liu H, et al. The 11β-hydroxysteroid dehydrogenase type 1 inhibitor protects against the insulin resistance and hepatic steatosis in db/db mice. European Journal of Pharmacology. 2016;788:140–51.

310. Feng L, Luo H, Xu Z, Yang Z, Du G, Zhang Y, et al. Bavachinin, as a novel natural pan-PPAR agonist, exhibits unique synergistic effects with synthetic PPAR-γ and PPAR-α agonists on carbohydrate and lipid metabolism in db/db and diet-induced obese mice. Diabetologia. 2016;59(6):1276–86.

311. Saito T, Hasegawa-Moriyama M, Yamada T, Kurimoto T, Inada E, Kanmura Y. Resolution of inflammation by resolvin D1 is essential for peroxisome proliferator-Activated receptor-γ-mediated analgesia during postincisional pain development in type 2 diabetes. Anesthesiology. 2015;123(6):1420–34.

312. Park MH, Nam YH, Han JS. Sargassum coreanum extract alleviates hyperglycemia and improves insulin resistance in db/db diabetic mice. Nutrition Research and Practice. 2015;9(5):472–9.

313. Park MH, Han J-S. *Padina arborescens* Ameliorates Hyperglycemia and Dyslipidemia in C57BL/KsJ-*db/db* Mice, a Model of Type 2 Diabetes Mellitus. Journal of Medicinal Food. 2015;18(10):1088–94.

314. Gim HJ, Li H, Jeong JH, Lee SJ, Sung M-K, Song M-Y, et al. Design, synthesis, and biological evaluation of a series of alkoxy-3-indolylacetic acids as peroxisome proliferator-activated receptor gamma/delta agonists. Bioorganic & medicinal chemistry. 2015;23(13):3322–36.

315. Sharma BR, Kim HJ, Rhyu DY. Caulerpa lentillifera extract ameliorates insulin resistance and regulates glucose metabolism in C57BL/KsJ-db/db mice via PI3K/AKT signaling pathway in myocytes. Journal of translational medicine. 2015;13(62):1–10.

316. Kotlinowski J, Grochot-Przeczek A, Taha H, Kozakowska M, Pilecki B, Skrzypek K, et al. PPARγ activation but not PPARγ haplodeficiency affects proangiogenic potential of endothelial cells and bone marrow-derived progenitors. Cardiovascular Diabetology. 2014 Nov;13(1):150.

317. Hanf R, Millatt LJ, Cariou B, Noel B, Rigou G, Delataille P, et al. The dual peroxisome proliferator-activated receptor alpha/delta agonist GFT505 exerts anti-diabetic effects in db/db mice without peroxisome proliferator-activated receptor gamma-associated adverse cardiac effects. Diabetes & vascular disease research : official journal of the International Society of Diabetes and Vascular Disease. 2014;11(6):440–7.

318. Mishra P, Singh SV, Verma AK, Srivastava P, Sultana S, Rath SK. Rosiglitazone induces cardiotoxicity by accelerated apoptosis. Cardiovascular Toxicology. 2014;14(2):99–119.

319. Yang H-I, Kim WS, Kim D-H, Kang JS. Histopathological Evaluation of Heart Toxicity of a Novel Selective PPAR-gamma Agonists CKD-501 in db/db Mice. Biomolecules & therapeutics. 2013;21(1):84–8.

320. Ye H, Zhang HJ, Xu A, Hoo RLC. Resistin Production from Adipose Tissue Is Decreased in db/db Obese Mice, and Is Reversed by Rosiglitazone. PLoS ONE. 2013;8(6):1–8.

321. Chodavarapu H, Grobe N, Somineni HK, Salem ESB, Madhu M, Elased KM. Rosiglitazone Treatment of Type 2 Diabetic db/db Mice Attenuates Urinary Albumin and Angiotensin Converting Enzyme 2 Excretion. PLoS ONE. 2013;8(4):1–12.

322. Zhou Y, Jia S, Wang C, Chen Z, Chi Y, Li J, et al. FAM3A is a target gene of peroxisome proliferator-activated receptor gamma. Biochimica et biophysica acta. 2013;1830(8):4160–70.

323. Wein S, Schrader E, Rimbach G, Wolffram S. Oral green tea catechins transiently lower plasma glucose concentrations in female db/db mice. Journal of medicinal food. 2013;16(4):312–7.

324. Shi D, Guo S, Jiang B, Guo C, Wang T, Zhang L, et al. HPN, a synthetic analogue of bromophenol from red alga Rhodomela confervoides: Synthesis and anti-diabetic effects in C57BL/KsJ-db/db mice. Marine Drugs. 2013;11(2):350–62.

325. He BK, Ning ZQ, Li ZB, Shan S, Pan DS, Ko BCB, et al. In vitro and in vivo characterizations of chiglitazar, a newly identified PPAR pan-agonist. PPAR Research. 2012;2012:1–13.

326. Schrader E, Wein S, Kristiansen K, Christensen LP, Rimbach G, Wolffram S. Plant Extracts of Winter Savory, Purple Coneflower, Buckwheat and Black Elder Activate PPAR-γ in COS-1 Cells but do not Lower Blood Glucose in Db/db Mice In vivo. Plant Foods for Human Nutrition. 2012;67(4):377–83.

327. Do GM, Jung UJ, Park HJ, Kwon EY, Jeon SM, Mcgregor RA, et al. Resveratrol ameliorates diabetes-related metabolic changes via activation of AMP-activated protein kinase and its downstream targets in db/db mice. Molecular Nutrition and Food Research. 2012;56(8):1282–91.

328. Park HJ, Jung UJ, Cho SJ, Jung HK, Shim S, Choi MS. Citrus unshiu peel extract ameliorates hyperglycemia and hepatic steatosis by altering inflammation and hepatic glucose- and lipid-regulating enzymes in db/db mice. Journal of Nutritional Biochemistry. 2013;24(2):419–27.

329. Tan Y, Muise ES, Dai H, Raubertas R, Wong KK, Thompson GM, et al. Novel Transcriptome Profiling Analyses Demonstrate that Selective PPARgamma Modulators Display Attenuated and Selective Gene Regulatory Activity in Comparison with PPARgamma Full Agonists. MolPharmacol. 2012;82(1):68–79.

330. Park CJ, Lee H-A, Han J-S. Jicama (Pachyrhizus erosus) extract increases insulin sensitivity and regulates hepatic glucose in C57BL/Ksj-db/db mice. Journal of Clinical Biochemistry and Nutrition. 2016;58(1):56–63.

331. Ortsater H, Grankvist N, Wolfram S, Kuehn N, Sjoholm A. Diet supplementation with green tea extract epigallocatechin gallate prevents progression to glucose intolerance in db/db mice. Nutr Metab (Lond). 2012;9:11.

332. Sharma a. N, Elased KM, Lucot JB. Rosiglitazone treatment reversed depression- but not psychosis-like behavior of db/db diabetic mice. Journal of Psychopharmacology. 2012;26(5):724–32.

333. Lee S-H, Min K-H, Han J-S, Lee D-H, Park D-B, Jung W-K, et al. Effects of brown alga, Ecklonia cava on glucose and lipid metabolism in C57BL/KsJ-db/db mice, a model of type 2 diabetes mellitus. Food and Chemical Toxicology. 2012;50(3–4):575–82.

334. Tang YH, Sun ZL, Fan MS, Li ZX, Huang CG. Anti-Diabetic Effects of TongGuanWan, a Chinese Traditional Herbal Formula, in C57BL/KsJ-db / db Mice. Planta Med. 2012;78(1):18–23.

335. Oishi K, Tomita T, Itoh N, Ohkura N. PPARγ activation induces acute PAI-1 gene expression in the liver but not in adipose tissues of diabetic model mice. Thrombosis Research. 2011;128(5):e81–5.

336. Keil S, Matter H, Schönafinger K, Glien M, Mathieu M, Marquette J-P, et al. Sulfonylthiadiazoles with an unusual binding mode as partial dual peroxisome proliferator-activated receptor (PPAR) γ/δ agonists with high potency and in vivo efficacy. ChemMedChem. 2011;6(4):633–53.

337. Le Bouter S, Rodriguez M, Guigal-Stephan N, Courtade-Gaïani S, Xuereb L, De Montrion C, et al. Coordinate transcriptomic and metabolomic effects of the insulin sensitizer rosiglitazone on fundamental metabolic pathways in liver, soleus muscle, and adipose tissue in diabetic db/db mice. PPAR Research. 2010;2010:1–17.

338. Yang KJ, Noh JR, Kim YH, Gang GT, Hwang JH, Yang SJ, et al. Differential modulatory effects of rosiglitazone and pioglitazone on white adipose tissue in db/db mice. Life Sciences. 2010;87(13–14):405–10.

339. Seto S, Okada K, Kiyota K, Isogai S, Iwago M, Shinozaki T, et al. Design, synthesis, and structure-activity relationship studies of novel 2,4,6-trisubstituted-5-pyrimidinecarboxylic acids as peroxisome proliferator-activated receptor γ (PPARγ) partial agonists with comparable antidiabetic efficacy to rosiglitazone. Journal of Medicinal Chemistry. 2010;53(13):5012–24.

340. Jung UJ, Baek N-I, Chung H-G, Jeong T-S, Lee KT, Lee M-K, et al. Antilipogenic and hypolipidemic effects of ethanol extracts from two variants of Artemisia princeps Pampanini in obese diabetic mice. Journal of medicinal food. 2009;12(6):1238–44.

341. Bak EJ, Park HG, Kim JM, Yoo Y-J, Cha J-H. Inhibitory effect of evodiamine alone and in combination with rosiglitazone on in vitro adipocyte differentiation and in vivo obesity related to diabetes. International journal of obesity (2005). 2010;34(2):250–60.

342. Cheng L, Han X, Shi Y. A regulatory role of LPCAT1 in the synthesis of inflammatory lipids, PAF and LPC, in the retina of diabetic mice. American journal of physiology Endocrinology and metabolism. 2009;297(6):E1276–82.

343. Au WS, Lu LW, Tam S, Ko OKH, Chow BKC, He ML, et al. Pluronic L-81 ameliorates diabetic symptoms in db/db mice through transcriptional regulation of microsomal triglyceride transfer protein. World Journal of Gastroenterology. 2009;15(24):2987–94.

344. Yun SI, Park HO, Kang JH. Effect of Lactobacillus gasseri BNR17 on blood glucose levels and body weight in a mouse model of type 2 diabetes. Journal of Applied Microbiology. 2009;107(5):1681–6.

345. Wu HS, Zhu DF, Zhou CX, Feng CR, Lou YJ, Yang B, et al. Insulin sensitizing activity of ethyl acetate fraction of Acorus calamus L. in vitro and in vivo. Journal of Ethnopharmacology. 2009;123(2):288–92.

346. Hui X, Zhu W, Wang Y, Lam KSL, Zhang J, Wu D, et al. Major urinary protein-1 increases energy expenditure and improves glucose intolerance through enhancing mitochondrial function in skeletal muscle of diabetic mice. Journal of Biological Chemistry. 2009;284(21):14050–7.

347. Park SY, Shin HK, Lee JH, Kim CD, Lee WS, Rhim BY, et al. Cilostazol Ameliorates Metabolic Abnormalities with Suppression of Proinflammatory Markers in a db/db Mouse Model of Type 2 Diabetes via Activation of Peroxisome Proliferator-Activated Receptor □ Transcription □. Pharmacology. 2009;329(2):571–9.

348. Yeo J, Kang Y-J, Jeon S-M, Jung UJ, Lee M-K, Song H, et al. Potential Hypoglycemic Effect of an Ethanol Extract of *Gynostemma pentaphyllum* in C57BL/KsJ- *db/db* Mice. Journal of Medicinal Food. 2008;11(4):709–16.

349. Wilson KD, Li Z, Wagner R, Yue P, Tsao P, Nestorova G, et al. Transcriptome alteration in the diabetic heart by rosiglitazone: Implications for cardiovascular mortality. PLoS ONE. 2008;3(7):1–11.

350. Kim D-J, Jeong Y-J, Kwon J-H, Moon K-D, Kim H-J, Jeon S-M, et al. Beneficial Effect of Chungkukjang on Regulating Blood Glucose and Pancreatic β-Cell Functions in C75BL/KsJ-db/db Mice. JOURNAL OF MEDICINAL FOOD J Med Food. 2008;11(2):215–23.

351. Chang CH, McNamara LA, Wu MS, Muise ES, Tan Y, Wood HB, et al. A novel selective peroxisome proliferator-activator receptor-γ modulator-SPPARγM5 improves insulin sensitivity with diminished adverse cardiovascular effects. European Journal of Pharmacology. 2008;584(1):192–201.

352. Yeo J, Lee YH, Jeon SM, Jung UJ, Lee MK, Jung YM, et al. Supplementation of a novel microbial biopolymer, PGB1, from new Enterobacter sp. BL-2 delays the deterioration of type 2 diabetic mice. Journal of Microbiology and Biotechnology. 2007;17(12):1983–90.

353. Jung UJ, Baek NI, Chung HG, Bang MH, Jeong TS, Tae Lee K, et al. Effects of the ethanol extract of the roots of Brassica rapa on glucose and lipid metabolism in C57BL/KsJ-db/db mice. Clinical Nutrition. 2008;27(1):158–67.

354. Yamanaka M, Itakura Y, Tsuchida a, Nakagawa T, Noguchi H, Taiji M. Comparison of the antidiabetic effects of brain-derived neurotrophic factor and thiazolidinediones in obese diabetic mice. Diabetes, obesity & metabolism. 2007;9(6):879–88.

355. Koh EH, Park JY, Park HS, Jeon MJ, Ryu JW, Kim M, et al. Essential role of mitochondrial function in adiponectin synthesis in adipocytes. Diabetes. 2007;56(12):2973–81.

356. How O-J, Larsen TS, Hafstad a D, Khalid A, Myhre ESP, Murray a J, et al. Rosiglitazone treatment improves cardiac efficiency in hearts from diabetic mice. Archives of physiology and biochemistry. 2007;113(4–5):211–20.

357. Jung UJ, Baek NI, Chung HG, Bang MH, Yoo JS, Jeong TS, et al. The anti-diabetic effects of ethanol extract from two variants of Artemisia princeps Pampanini in C57BL/KsJ-db/db mice. Food and Chemical Toxicology. 2007;45(10):2022–9.

358. Roy S, Khanna V, Mittra S, Dhar A, Singh S, Mahajan DC, et al. Combination of dipeptidylpeptidase IV inhibitor and low dose thiazolidinedione: Preclinical efficacy and safety in db/db mice. Life Sciences. 2007;81(1):72–9.

359. Okamoto Y, Higashiyama H, Rong JX, McVey MJ, Kinoshita M, Asano S, et al. Comparison of mitochondrial and macrophage content between subcutaneous and visceral fat in db/db mice. Experimental and Molecular Pathology. 2007;83(1):73–83.

360. Mittra S, Sangle G, Tandon R, Sharma S, Roy S, Khanna V, et al. Increase in weight induced by muraglitazar, a dual PPARalpha/gamma agonist, in db/db mice: adipogenesis/or oedema? Br J Pharmacol. 2007;150(4):480–7.

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

361. Han HO, Koh JS, Kim SH, Park OK, Kim KH, Jeon SK, et al. Design and synthesis of Oxime ethers of α-acyl-β-phenylbutanoic acids as PPAR α and -γ dual agonists. Bioorganic & medicinal chemistry letters. 2007;17(4):937–41.

362. Wang X, Hu Z, Hu J, Du J, Mitch WE. Insulin resistance accelerates muscle protein degradation: Activation of the ubiquitin-proteasome pathway by defects in muscle cell signaling. Endocrinology. 2006;147(9):4160–8.

363. Choo HJ, Kim JH, Kwon OB, Lee CS, Mun JY, Han SS, et al. Mitochondria are impaired in the adipocytes of type 2 diabetic mice. Diabetologia. 2006;49(4):784–91.

364. Loffler M, Bilban M, Reimers M, Waldhäusl W, Stulnig TM. Blood glucose-lowering nuclear receptor agonists only partially normalize hepatic gene expression in db/db mice. The Journal of pharmacology and experimental therapeutics. 2006;316(2):797–804.

365. Dropinski JF, Akiyama T, Einstein M, Habulihaz B, Doebber T, Berger JP, et al. Synthesis and biological activities of novel aryl indole-2-carboxylic acid analogs as PPARγ partial agonists. Bioorganic and Medicinal Chemistry Letters. 2005;15(22):5035–8.

366. Shi GQ, Dropinski JF, McKeever BM, Xu S, Becker JW, Berger JP, et al. Design and synthesis of alpha-aryloxyphenylacetic acid derivatives: a novel class of PPARalpha/gamma dual agonists with potent antihyperglycemic and lipid modulating activity. Journal of medicinal chemistry. 2005;48(13):4457–68.

367. Li DD, Chen JH, Chen Q, Li GW, Chen J, Yue JM, et al. Swietenia mahagony extract shows agonistic activity to PPARγ and gives ameliorative effects on diabetic db/db mice. Acta Pharmacologica Sinica. 2005;26(2):220–2.

368. Minoura H, Takeshita S, Ita M, Hirosumi J, Mabuchi M, Kawamura I, et al. Pharmacological characteristics of a novel nonthiazolidinedione insulin sensitizer, FK614. European Journal of Pharmacology. 2004;494(2–3):273–81.

369. Kast-Woelbern HR, Dana SL, Cesario RM, Sun L, De Grandpre LY, Brooks ME, et al. Rosiglitazone induction of Insig-1 in white adipose tissue reveals a novel interplay of peroxisome proliferator-activated receptor γ and sterol regulatory element-binding protein in the regulation of adipogenesis. Journal of Biological Chemistry. 2004;279(23):23908–15.

370. Arlt W, Neogi P, Gross C, Miller WL. Cinnamic acid based thiazolidinediones inhibit human P450c17 and 3 beta-hydroxysteroid dehydrogenase and improve insulin sensitivity independent of PPAR gamma agonist activity. Journal of Molecular Endocrinology. 2004;32:425–36.

371. Shen Q, Cline GW, Shulman GI, Leibowitz MD, Davies PJA. Effects of Rexinoids on Glucose Transport and Insulin-mediated Signaling in Skeletal Muscles of Diabetic (db/db) Mice. Journal of Biological Chemistry. 2004;279(19):19721–31.

372. Bagi Z, Koller A, Kaley G. PPARgamma activation, by reducing oxidative stress, increases NO bioavailability in coronary arterioles of mice with Type 2 diabetes. American journal of physiology Heart and circulatory physiology. 2004;286(2):H742–8.

373. Desai RC, Han W, Metzger EJ, Bergman JP, Gratale DF, MacNaul KL, et al. 5-Aryl thiazolidine-2,4-diones: Discovery of PPAR dual α/γ agonists as antidiabetic agents. Bioorganic and Medicinal Chemistry Letters. 2003;13(16):2795–8.

374. Misra P, Chakrabarti R, Vikramadithyan RK, Bolusu G, Juluri S, Hiriyan J, et al. PAT5A: a partial agonist of peroxisome proliferator-activated receptor gamma is a potent antidiabetic thiazolidinedione yet weakly adipogenic. The Journal of pharmacology and experimental therapeutics. 2003;306(2):763–71.

375. Koyama H, Boueres JK, Han W, Metzger EJ, Bergman JP, Gratale DF, et al. 5-Aryl thiazolidine-2,4-diones as selective PPARγ agonists. Bioorganic and Medicinal Chemistry Letters. 2003;13(10):1801–4.

376. Edwards D. Non-linear normalization and background correction in one-channel cDNA microarray studies. Bioinformatics. 2003;19(7):825–33.

377. Adams AD, Yuen W, Hu Z, Santini C, Jones AB, MacNaul KL, et al. Amphipathic 3-phenyl-7-propylbenzisoxazoles; human PPaR γ, δ and α agonists. Bioorganic and Medicinal Chemistry Letters. 2003;13(5):931–5.

378. Vikramadithyan RK, Hiriyan J, Suresh J, Gershome C, Babu RK, Misra P, et al. DRF 2655: a unique molecule that reduces body weight and ameliorates metabolic abnormalities. Obesity research. 2003;11(2):293–303.

379. Chakrabarti R, Vikramadithyan RK, Prem Kumar M, Kumar SKB, Mamidi NVS, Misra P, et al. PMT13, a pyrimidone analogue of thiazolidinedione improves insulin resistance-associated disorders in animal models of type 2 diabetes. Diabetes, Obesity and Metabolism. 2002;4(5):319–28.

380. Hori H, Sasaoka T, Ishihara H, Wada T, Murakami S, Ishiki M, et al. Association of SH2-Containing Inositol Phosphatase2 With the Insulin Resistance of Diabetic db/db Mice. Diabetes. 2002;51(18):2387–94.

381. Madhavan GR, Chakrabarti R, Vikramadithyan RK, Mamidi RNVS, Balraju V, Rajesh BM, et al. Synthesis and biological activity of novel pyrimidinone containing thiazolidinedione derivatives. Bioorganic and Medicinal Chemistry. 2002;10(8):2671–80.

382. Gerhold DL, Liu F, Jiang G, Li Z, Xu J, Lu M, et al. Gene expression profile of adipocyte differentiation and its regulation by peroxisome proliferator-activated receptor-gamma agonists. Endocrinology. 2002;143(6):2106–18.

383. Etgen GJ, Oldham BA, Johnson WT, Broderick CL, Montrose CR, Brozinick JT, et al. A tailored therapy for the metabolic syndrome: The dual peroxisome proliferator-activated receptor-alpha/gamma agonist LY465608 ameliorates insulin resistance and diabetic hyperglycemia while improving cardiovascular risk factors in preclinical models. Diabetes. 2002;51(4):1083–7.

384. Clapham JC, Coulthard VH, Moore GB. Concordant mRNA expression of UCP-3, but not UCP-2, with mitochondrial thioesterase-1 in brown adipose tissue and skeletal muscle in db/db diabetic mice. Biochemical and biophysical research communications. 2001;287(5):1058–62.

385. Moore GB, Chapman H, Holder JC, Lister CA, Piercy V, Smith SA, et al. Differential regulation of adipocytokine mRNAs by rosiglitazone in db/db mice. Biochemical and biophysical research communications. 2001;286(4):735–41.

386. Lohray BB, Lohray VB, Bajji AC, Kalchar S, Poondra RR, Padakanti S, et al. (-)3-[4-[2-(phenoxazin-10-yl)ethoxy]phenyl]-2-ethoxypropanoic acid [(-)DRF 2725]: A dual PPAR agonist with potent antihyperglycemic and lipid modulating activity. Journal of Medicinal Chemistry. 2001;44(16):2675–8.

387. Vikramadithyan RK, Chakrabarti R, Misra P, Premkumar M, Kumar SKB, Rao CS, et al. Euglycemic and hypolipidemic activity of PAT5A: A unique thiazolidinedione with weak peroxisome proliferator activated receptor gamma activity. Metabolism: Clinical and Experimental. 2000;49(11):1417–24.

388. Chaput E, Saladin R, Silvestre M, Edgar AD. Fenofibrate and rosiglitazone lower serum triglycerides with opposing effects on body weight. Biochemical and biophysical research communications. 2000;271(2):445–50.

389. Lohray BB, Bhushan V, Reddy AS, Rao PB, Reddy NJ, Harikishore P, et al. Novel euglycemic and hypolipidemic agents. 4. Pyridyl- and quinolinyl- containing thiazolidinediones. Journal of Medicinal Chemistry. 1999;42(14):2569–81.

390. Burris TP, Pelton PD, Zhou L, Osborne MC, Cryan E, Demarest KT. A novel method for analysis of nuclear receptor function at natural promoters: peroxisome proliferator-activated receptor gamma agonist actions on aP2 gene expression detected using branched DNA messenger RNA quantitation. Molecular endocrinology (Baltimore, Md). 1999;13(3):410–7.

391. Lohray BB, Bhushan V, Rao BP, Madhavan GR, Murali N, Rao KN, et al. Novel euglycemic and hypolipidemic agents. Journal of Medicinal Chemistry. 1998;41(10):1619–30.

392. Connor SC, Hughes MG, Moore G, Lister CA, Smith SA. Antidiabetic efficacy of BRL 49653, a potent orally active insulin sensitizing agent, assessed in the C57BL/KsJ db/db diabetic mouse by non-invasive 1H NMR studies of urine. The Journal of pharmacy and pharmacology. 1997;49(3):336–44.

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

393. Yoshida S, Ohishi T, Matsui T, Tanaka H, Oshima H, Yonetoku Y, et al. Novel GPR119 agonist AS1535907 contributes to first-phase insulin secretion in rat perfused pancreas and diabetic db/db mice. Biochemical and Biophysical Research Communications. 2010;402(2):280–5.

394. Carroll MJ, Lister CA, Sennitt M V., Stewart-Long N, Cawthorne MA. Improved glycemic control in C57Bl/KsJ (db/db) mice after treatment with the thermogenic β-adrenoceptor agonist, BRL 26830. Diabetes. 1985;34(11):1198–204.

395. Chakrabarti R, Damarla RKB, Mullangi R, Sharma VM, Vikramadithyan RK, Rajagopalan R. Insulin sensitizing property of Indigofera mysorensis extract. Journal of Ethnopharmacology. 2006;105(1–2):102–6.

396. Han X, Deng Y, Yu J, Sun Y, Ren G, Cai J, et al. Acarbose Accelerates Wound Healing via Akt/eNOS Signaling in db/db Mice. Oxidative Medicine and Cellular Longevity. 2017;2017:1–11.

397. Lee KH, Ha KS, Jo SH, Lee CM, Kim YC, Chung KH, et al. Effect of long-term dietary Arginyl-Fructose (AF) on hyperglycemia and HbA1c in diabetic db/db mice. International Journal of Molecular Sciences. 2014;15(5):8352–9.

398. Kim J-G, Jo S-H, Ha K-S, Kim S-C, Kim Y-C, Apostolidis E, et al. Effect of long-term supplementation of low molecular weight chitosan oligosaccharide (GO2KA1) on fasting blood glucose and HbA1c in db/db mice model and elucidation of mechanism of action. BMC complementary and alternative medicine. 2014;14(1):272.

399. Kim J-H, Kang M-J, Choi H-N, Jeong S-M, Lee Y-M, Kim J-I. Quercetin attenuates fasting and postprandial hyperglycemia in animal models of diabetes mellitus. Nutrition research and practice. 2011;5(2):107–11.

400. Kang MJ, Kim JH, Choi HN, Kim MJ, Han JH, Lee JH, et al. Hypoglycemic effects of Welsh onion in an animal model of diabetes mellitus. Nutrition Research and Practice. 2010;4(6):486–91.

401. Lee SM. The effect of chronic alpha-glycosidase inhibition on diabetic nephropathy in the db/db mouse. Diabetes. 1982;31(3):249–54.

402. Chang MS, Oh MS, Kim DR, Jung KJ, Park S, Choi SB, et al. Effects of Okchun-San, a herbal formulation, on blood glucose levels and body weight in a model of Type 2 diabetes. Journal of Ethnopharmacology. 2006;103(3):491–5.

Appendix E – Validation Duchenne Muscular Dystrophy (DMD)

MODEL NAME	C57Bl10scsn-Dmdmdx mouse (mdx mouse)
INDICATION	Duchenne Muscular Dystrophy
VALIDATION DATE	19.09.2016
TOTAL SUBSECTIONS	33
TOTAL SCORE	67.78
UNCERTAINTY FACTOR (%)	6.06
VALIDATION LEVEL (%)	Highly Validated (93.9)
HISTORICAL BACKGROUND	

The muscular dystrophy X-chromosome linked (mdx) mouse is the most widely used animal model to study Duchenne Muscular Dystrophy (DMD) [6]. It was first described by Bulfield et al in the 80's, after inbreeding C57BL/10ScSn mice for five generations and noticing increased creatine kinase (CK) and pyruvate kinase (PK). Histopathological muscle damage similar to DMD patients was also observed [7]. Later this spontaneous mutation was characterised by Sicinski et al as being a nonsense point mutation (C-to-T transition) in exon 23, which leads to failure in expressing full-length dystrophin [8].

1. EPIDEMIOLOGICAL VALIDATION	
1.1 Is the model able to simulate the disease in the relevant sexes?	Score
Yes, completely. **Remarks**: most literature published on mdx mice reports the use of male animals. DMD occurs almost exclusively in males with an incidence of 1 every 3,500 to 6,291 births [9,10].	6.25
1.2 Is the model able to simulate the disease in the relevant age groups (juvenile, adult or ageing)?	Score
Yes, partially. **Remarks**: DMD is a disease caused by a genetic defect in the dystrophin coding gene. Humans are affected during infancy, with clinical presentation at around 2.5 years. Mdx mice present pathological signs at 3 weeks, but clinical symptoms appear only in late adulthood (12 to 15 months). Hence, there is only a limited period the disease overlaps in both species [11–16].	3.13

2. SYMPTOMATOLOGY AND NATURAL HISTORY (SNH) VALIDATION	
2.1 Is the model able to replicate the symptoms and co-morbidities commonly present in this disease? If so, which ones?	Score
Yes, partially.	1.61
2.1.1 Symptoms modelled	

Reduced grown-up body weight (BW): BW is reduced in mdx mice when compared to control or wild-type mice [14,17].

ECG abnormality: similar to humans, ECG changes such as sinus tachycardia, shortened PR interval, prolonged QT interval, deep Q wave, and polyphasic R' wave are observed in old mdx mice [14].

Cardiomyopathy: fibrosis, calcification and enlarged ventricular chamber are also present in mdx mice [7,14].

2.1.2 Symptoms partially modelled	
Reduced lifespan: approximately 18% reduced lifespan when compared to wild-type mice [18]. In humans, the reduction is more significant at around 75% [6,13].	
Muscle wasting: it is comparable to DMD boys only regarding diaphragm degeneration [7,18,19].	
Cognitive and CNS defects: mdx mice show a deficit in cognitive deficit flexibility, long-term spatial and recognition memory, although somewhat milder than in boys in with DMD and with no short-term memory involvement [20–26].	

2.1.3 Symptoms not modelled	
Loss of ambulation: hardly ever present in mdx mice while for humans it happens between 10 and 15 years – the latter being for patients with less damaging exon skipping and treatment with glucocorticoids [6,12,13,17].	

2.2 Is the natural history of the disease similar to humans regarding:	

2.2.1 Time to onset;	Score
No. **Remarks**: although histological marks of disease can be detected early (at 3 weeks is the start of limb muscle degeneration), real impairment (muscle wasting, scoliosis and heart failure) do not occur until mice are 15 months or older – equivalent to around 26 years in humans [7,14–17,27–29]. In humans, first symptoms start around 2.5 years, with loss of ambulation around 10 to 15 years [11–13].	0

2.2.2 Disease progression;	Score
Yes, partially. **Remarks**: mild progression, as most of the limb muscles maintain hypertrophy although with a loss in specific force and normalised power unlike DMD boys, whose muscles' degenerative processes largely outdistance the regenerative, rapidly causing muscle atrophy [6,18,30]. Although histological marks of disease can be detected early (at 3 weeks is the start of limb muscle degeneration; at 5 weeks is the start of muscle necrosis; at 9 weeks (adult mice): all muscles are affected), real impairment (muscle wasting, scoliosis and heart failure) do not occur until mice are 15 months or older [7,14–17,27,28]. At ~12 to 15 months old mdx mice show abnormal histopathology, echocardiography and ECDG and at 21 months old (eq. to ~53 years in humans) mdx mice show signs of cardiomyopathy [14]. In humans, first symptoms start around 2.5 years, with loss of ambulation around 10 to 15 years [11–13].	1.25

2.2.3 Duration of symptoms;	Score
Yes, completely. **Remarks**: symptoms in DMD are lifelong due to permanent muscle degeneration. In mdx mice, once symptoms get worse, there is also permanent damage that leads to reduced lifespan [18].	2.50

2.2.4 Severity.	Score
Yes, partially. **Remarks**: mdx mice do not show the same level of muscle degeneration as seen in humans, except for the diaphragm which shows progressive degeneration evident at 6 months of age with great variation in myofibre size, architecture and continuous necrosis and proliferation of connective tissue [7,18,19,30].	1.25

3. GENETIC VALIDATION	
3.1 Does this species also have orthologous genes and/or proteins involved in the human disease?	Score
Yes. **Remarks**: DMD is caused by the lack of dystrophin, thus the dystrophin gene alongside with utrophin (a functional and structural analogue) have been included in this section [31]. The mouse has orthologous genes for both of them [32,33].	4.17
3.1.1 Dystrophin	
The human dystrophin gene (Gene ID: 1756) is the largest gene found in humans measuring a total of 2.4 Mb located at Xp21.2-p21.1 [34]. Dystrophin forms parts of the dystrophin-glycoprotein complex (DGC), responsible for connecting the inner cytoskeleton and the extracellular matrix [34]. In mice, the dystrophin gene (Gene ID: 13405) is located at chromosome X C1; X 38.38 cM [32].	
3.1.2 Utrophin	
The human utrophin gene (Gene ID: 7402) is located in humans at 6q24.2 [35]. Utrophin is a structural and functional analogue to dystrophin, being also present at neuromuscular synapses and myotendinous junctions [35]. Utrophin upregulation is thought to be a compensation mechanism partially responsible for the milder phenotype in mdx mice [35]. In mice, the utrophin gene (Gene ID: 22288) is located at chromosome 10 A1-A2; 10 3.77 cM [33].	
3.2 If so, are the relevant genetic mutations or alterations also present in the orthologous genes/proteins?	Score
Yes, partially. **Remarks**: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X [31]. In that sense, DMD being an X-linked inherited disorder, the mdx mouse has the same aetiology (spontaneous mutations). Nevertheless, mdx mice have a nonsense point mutation (C-to-T transition) in exon 23 that aborts full-length dystrophin expression [8]. This is not the case for human patients who have other alterations that lead to dystrophin deficiency.	2.09
3.3 If so, is the expression of such orthologous genes and/or proteins similar to the human condition?	Score
Yes, completely.	4.16
3.3.1 Dystrophin	
Yes, completely. **Remarks**: like DMD patients, mdx mice are dystrophin-deficient, not expressing full-length dystrophin [8].	
3.3.2 Utrophin	
Yes, completely. **Remarks**: utrophin upregulation (a compensation mechanism partially responsible for the milder phenotype in mdx mice) is reported in mdx mice and humans [14,36,37].	

4. BIOCHEMICAL VALIDATION	
4.1 If there are known pharmacodynamic (PD) biomarkers related to the pathophysiology of the disease, are they also present in the model?	Score
Yes, completely. **Remarks**: high levels of creatine kinase (CK) – a marker of muscle damage, with the muscle-type variants being predominant in plasma are present in mdx mice [7,38]. Pyruvate kinase (PK) is also increased, although it has secondary importance and it is a less reliable marker of muscle damage than CK, being included here only for information purposes [38].	3.13
4.2 Do these PD biomarkers behave similarly to humans'?	Score
Yes, completely. **Remarks**: both PK and CK are increased in mdx mice and humans [7,38].	3.13
4.3 If there are known prognostic biomarkers related to the pathophysiology of the disease, are they also present in the model?	Score
N/A. **Remarks**: currently, DMD has no prognostic biomarker that is validated and widely used in clinical practice. Many biomarkers are still under exploratory phase. A list of biomarkers being explored can be found in the supplementary material from Guiraud et al, 2015 and in the paper from Hathout et al, 2016 [39,40].	-
4.4 Do these prognostic biomarkers behave similarly to humans'?	Score
N/A.	-

5. AETIOLOGICAL VALIDATION	
5.1 Is the aetiology of the disease similar to humans'?	Score
Yes, partially. **Remarks**: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X [31]. In that sense, DMD being an X-linked inherited disorder, the mdx mouse has the same aetiology (spontaneous mutations). Nevertheless, mdx mice have a nonsense point mutation (C-to-T transition) in exon 23 that aborts full-length dystrophin expression [8]. This is not the case for human patients who have other alterations that lead to dystrophin deficiency.	6.25

6. HISTOLOGICAL VALIDATION	
6.1 Do the histopathological structures in relevant tissues resemble the ones found in humans?	Score
Yes, partially.	4.17
6.1.1 Histopathological features modelled	
Limb muscle fibrosis: like in DMD boys, it is significantly increased in the mdx mouse muscles [16,17,27].	
6.1.2 Histopathological features not modelled	
Muscle regeneration: in mdx mice, muscle regeneration is significantly higher than in humans until old age [17,27,28]. **Adipose tissue**: scarcely developed in mdx muscle, unlike in DMD patients [17,28]. **Remarks**: acute necrosis is present in mdx mice muscles but not in humans [17].	-

7. PHARMACOLOGICAL VALIDATION
7.1 Are effective drugs in humans also effective in this model?   Score
3.61

7.1.1 Glucocorticosteroids

7.1.1.1 Prednisone [41–55]

Methodology: studies were searched on 13/09/2016 on PubMed with the string 'mdx AND (Prednisone[mesh] OR prednisone OR Apo-Prednisone OR Cortan OR Cortancyl OR Cutason OR Dacortin OR Decortin OR Decortisyl OR Dehydrocortisone OR Deltasone OR Encorton OR Encortone OR Enkortolon OR Kortancyl OR "Liquid Pred" OR Meticorten OR Orasone OR Panafcort OR Panasol OR Predni Tablinen OR Prednidib OR Predniment OR Prednison Acsis OR "Prednison Galen" OR "Prednison Hexal" OR Pronisone OR Rectodelt OR Sone OR Sterapred OR Ultracorten OR Winpred OR delta-Cortisone)', yielding 27 results; and on Embase with the string '("Duchenne muscular dystrophic mouse" OR "Duchenne muscular dystrophy mouse" OR "MDX mouse" OR "mice, inbred mdx" OR "X chromosome-linked muscular dystrophy mouse" OR "X-linked muscular dystrophic mouse" OR "X-linked muscular dystrophy mouse") AND (prednisone OR "1, 2 dehydrocortisone" OR "17, 21 dihydroxypregna 1, 4 diene 3, 11, 20 trione" OR ancortone OR apo-prednisone OR biocortone OR colisone OR cortan OR cortidelt OR cortiprex OR cutason OR dacorten OR "de cortisyl" OR decortancyl OR decortin OR "decortin e merck" OR decortine OR decortisyl OR dehydrocortisone OR dekortin OR delitisone OR "dellacort a" OR "delta 1 dehydrocortisone" OR "delta cortelan" OR "delta cortisone" OR "delta dome" OR "delta e" OR "delta prenovis" OR delta-dome OR deltacorten OR deltacortene OR deltacortisone OR deltacortone OR deltasone OR deltison OR deltisona OR deltra OR "di adreson" OR di-adreson OR diadreson OR drazone OR encorton OR encortone OR enkorton OR fernisone OR hostacortin OR insone OR "liquid pred" OR lodotra OR me-korti OR meprison OR metacortandracin OR meticorten OR meticortine OR nisona OR "nsc 10023" OR nsc10023 OR orasone OR orisane OR panafcort OR paracort OR pehacort OR precort OR precortal OR prednicen-m OR prednicorm OR prednicot OR prednidib OR prednison OR "prednisone alcohol" OR "prednisone intensol" OR "prednisone test" OR prednitone OR "pregna 1, 4 diene 3, 11, 20 trione 17, 21 diol" OR pronison OR pronisone OR pronizone OR pulmison OR rayos OR rectodelt OR servisone OR steerometz OR sterapred OR "sterapred ds" OR ultracorten OR urtilone OR winpred), yielding 34 results; which were screened by title and abstract. Of these, fifteen (15) articles were included in this section.

Results: a significant part of the studies shows a positive effect prednisone in histological (e.g. reduction of the percentage of fibres with centrally located nuclei, inflammatory cells infiltration etc.) and functional outcomes (e.g. fore- and hindlimb grip strengths, respiratory function and motor coordination), although in some studies prednisone did not show any effects on functional outcomes (e.g. total distance in open field tests, fore- and hindlimb grip strengths). However, there are also reports of an increase in heart muscle fibrosis. Seven (7) articles were classified as category I and eight (8) were classified as category II. Of the latter, six (6) studies are in line with clinical findings.

Reporting Quality		Risk of Bias	
Parameter (N = 15)	Y (%)	Parameter (N = 15)	Y/U (%)
Type of Facility	13.3	Allocation Concealment	0/100
Type of Cage or Housing	20.0	Blinded Outcome Assessment	6.7/93.3
Bedding Material	0.0	Blinded Operations	0/100
N Cage Companions	20.0	Random Cage Allocation	0/100
Breeding Programme	53.3	Random Outcome Assessment	0/100
Light/Dark Cycle	46.7	Sequence Generation	0/93.3
Temperature and Humidity	0.0	Baseline Characteristics	6.7/3.3
Quality of the Water (fish)	-	Incomplete Outcome Data	13.3/60.0
Type of Food	6.7	Selective Outcome Reporting	100/0
Access to Food and Water	53.3	Other	66.7/0
Environmental Enrichment	0.0		
Any Blinding	66.7		
Any Randomisation	40.0		
Sample Size	100		
Sample Size Calculation	13.3		
Acclimatisation	33.3		
Sex Disclosed	53.3		
Male/Female/Both (N = 8)	62.5/25.0/12.5		
Background Control	-		
Background Model	-		

7.1.1.2 Deflazacort [45,46,48,49,56–63]

Methodology: studies were searched on 13/09/2016 on PubMed with the string 'mdx AND deflazacort', yielding 15 results; and on Embase with the string '("Duchenne muscular dystrophic mouse" OR "Duchenne muscular dystrophy mouse" OR "MDX mouse" OR "mice, inbred mdx" OR "X chromosome-linked muscular dystrophy mouse" OR "X-linked muscular dystrophic mouse" OR "X-linked muscular dystrophy mouse") AND deflazacort OR "9 defluorofluazacort" OR azacort OR calcort OR deflan OR defluorofluazacort OR "dl 458" OR "dl 458 it" OR "dl 458it" OR "dl 5458" OR dl458 OR dl458it OR dl5458 OR emflaza OR flantadin OR "fluazacort, difluoro" OR rosilar, yielding 25 results; which were screened by title and abstract. Of these, twelve (12) articles were included in this section.

Results: all animal studies show positive effects on histological outcomes (reduction of the percentage of fibres with centrally located nuclei, on the prevalence of dystrophic lesions, fibre size variability, the density of inflammatory cells and area). Of the few which investigated functional outcomes (e.g. increase in total run distance and hindlimb grip strengths and reduction in muscle fatigue), all demonstrated benefits on muscle function although no effect was reported on exploratory behaviour. Nine (9) articles were classified as category I and three (3) were classified as category II. Of the latter, all studies are in line with clinical findings.

Reporting Quality		Risk of Bias	
Parameter (N = 12)	Y (%)	Parameter (N = 12)	Y/U (%)
Type of Facility	0.0	Allocation Concealment	0/100
Type of Cage or Housing	0.0	Blinded Outcome Assessment	0/100
Bedding Material	0.0	Blinded Operations	0/100
N Cage Companions	0.0	Random Cage Allocation	0/100
Breeding Programme	41.7	Random Outcome Assessment	0/100
Light/Dark Cycle	0.0	Sequence Generation	0/100
Temperature and Humidity	0.0	Baseline Characteristics	0/100
Quality of the Water (fish)	-	Incomplete Outcome Data	33.3/58.3
Type of Food	0.0	Selective Outcome Reporting	100/0
Access to Food and Water	8.3	Other	83.3/0
Environmental Enrichment	0.0		
Any Blinding	58.3		
Any Randomisation	41.7		
Sample Size	91.7		
Sample Size Calculation	0.0		
Acclimatisation	0.0		
Sex Disclosed	25.0		
Male/Female/Both (N = 3)	33.3/0/66.7		
Background Control	-		
Background Model	-		

7.1.2 Read-through compound: Ataluren [64]

Methodology: studies were searched on 13/09/2016 on PubMed with the string 'mdx AND (ptc124 OR ataluren)', yielding 7 results; and on Embase with the string '("duchenne muscular dystrophic mouse" OR "duchenne muscular dystrophy mouse" OR "mdx mouse" OR "mice, inbred mdx" OR "x chromosome-linked muscular dystrophy mouse" OR "x-linked muscular dystrophic mouse" OR "x-linked muscular dystrophy mouse") AND ("3 [5 (2 fluorophenyl) 1, 2, 4 oxadiazol 3 yl] benzoic acid" OR "3 [5 (2 fluorophenyl) [1, 2, 4] oxadiazol 3 yl] benzoic acid" OR "ataluren sodium" OR "ptc 124" OR ptc124 OR translarna)', yielding 11 results; which were screened by title and abstract. Of these, one (1) article was included in this section.

Results: the only study show dystrophin detection in all muscles examined, including restoration of the dystrophin-glycoprotein complex (DCG), reduction of CK levels (linked to muscle damage), increase in average force and hanging time (4-limb hanging test) and reduction in functional strength deficit. This article was classified as category II. This study is in line with clinical findings.

Reporting Quality		Risk of Bias	
Parameter (N = 1)	Y (%)	Parameter (N = 1)	Y/U (%)
Type of Facility	0.0	Allocation Concealment	0/100
Type of Cage or Housing	0.0	Blinded Outcome Assessment	0/100
Bedding Material	0.0	Blinded Operations	0/100
N Cage Companions	0.0	Random Cage Allocation	0/100
Breeding Programme	100	Random Outcome Assessment	0/100
Light/Dark Cycle	0.0	Sequence Generation	0/100
Temperature and Humidity	0.0	Baseline Characteristics	0/100
Quality of the Water (fish)	-	Incomplete Outcome Data	0/100
Type of Food	0.0	Selective Outcome Reporting	0/100
Access to Food and Water	0.0	Other	100/0
Environmental Enrichment	0.0		
Any Blinding	0.0		
Any Randomisation	100		
Sample Size	0.0		
Sample Size Calculation	0.0		
Acclimatisation	0.0		
Sex Disclosed	100		
Male/Female/Both (N = 1)	100/0/0		
Background Control	-		
Background Model	-		

7.1.3 Exon skipping: Eteplirsen [65–69]

Methodology: As eteplirsen is designed to skip exon 51 in humans, it cannot itself be tested in mdx mice, which carry a mutation in exon 23. Thus, we also searched for AVI 4225, which is the mouse homologous compound of eteplirsen. Studies were searched on 19/09/2016 on PubMed with the string 'mdx AND '(eteplirsen OR avi-4658 OR "avi 4658" OR AVI-4225 OR "AVI 4225")', yielding 2 results; and on Embase with the string '("duchenne muscular dystrophic mouse" OR "duchenne muscular dystrophy mouse" OR "mdx mouse" OR "mice, inbred mdx" OR "x chromosome-linked muscular dystrophy mouse" OR "x-linked muscular dystrophic mouse" OR "x-linked muscular dystrophy mouse") AND (eteplirsen OR "avi 4658" OR avi4658 OR "avi 4225" OR avi-4225 OR etiplersen OR etiplirsen OR exondys OR "exondys 51")', yielding 7 results; which were screened by title and abstract. Of these, one article was included in this section. Moreover, when looking at FDA's Pharmacology Review of eteplirsen, four (4) studies are cited on the mdx mouse with eteplirsen's mouse homologous compound AVI-4225 [70]. These four (4) studies were also included in this section, leading to a total of five (5) studies.

Results: in general, all studies showed increased dystrophin expression in muscle. Histological measures generally improved, including the diaphragm and cardiac function. Muscle strength was also reported to have improved in functional measures (e.g. grip or tetanic strength). All articles but one were classified as category II. All studies are in line with clinical findings.

Reporting Quality		Risk of Bias	
Parameter (N = 5)	Y (%)	Parameter (N = 5)	Y/U (%)
Type of Facility	0.0	Allocation Concealment	0/100
Type of Cage or Housing	20.0	Blinded Outcome Assessment	0/100
Bedding Material	0.0	Blinded Operations	0/100
N Cage Companions	20.0	Random Cage Allocation	0/100
Breeding Programme	60.0	Random Outcome Assessment	0/100
Light/Dark Cycle	20.0	Sequence Generation	0/100
Temperature and Humidity	20.0	Baseline Characteristics	0/100
Quality of the Water (fish)	-	Incomplete Outcome Data	20.0/80.0
Type of Food	0.0	Selective Outcome Reporting	80.0/20.0
Access to Food and Water	40.0	Other	80.0/20.0
Environmental Enrichment	0.0		
Any Blinding	0.0		
Any Randomisation	20.0		
Sample Size	80.0		
Sample Size Calculation	0.0		
Acclimatisation	0.0		
Sex Disclosed	25.0		
Male/Female/Both (N = 2)	50/50/0		
Background Control	-		
Background Model	-		

7.2 Are ineffective drugs in humans also ineffective in this model?	Score
Two identified drug classes comprise drugs that were tested in humans and failed due to the lack of efficacy: aminopeptidase and leucine peptidase inhibitors; and myostatin inhibitors.	0.21

7.2.1 Aminopeptidase and leucine peptidase inhibitors: Bestatin [71]

Methodology: studies were searched on 19/09/2016 on PubMed with the string ' mdx AND (Ubenimex[mesh] OR ubenimex OR "uben 3-amino-2-hydroxy-4-phenylbutyryl-L-leucine" OR bestatin OR "bestatin, (D-Leu)-(R-(R*,R*))-isomer" OR "bestatin, (D-Leu)-(R-(R*,S*))-isomer" OR "bestatin, (D-Leu)-(S-(R*,S*))-isomer" OR "bestatin, (L-Leu)-(R-(R*,R*))-isomer OR bestatin, (L-Leu)-(R-(R*,S*))-isomer" OR "bestatin, (L-Leu)-(S-(R*,R*))-isomer" OR "bestatin, (L-Leu)-(S-(R*,S*))-isomer, hydrochloride")', yielding 1 result; and on Embase with the string '("duchenne muscular dystrophic mouse" OR "duchenne muscular dystrophy mouse" OR "mdx mouse" OR "mice, inbred mdx" OR "x chromosome-linked muscular dystrophy mouse" OR "x-linked muscular dystrophic mouse" OR "x-linked muscular dystrophy mouse") AND (bestatin OR "(3 amino 2 hydroxy 4 phenylbutanoyl) leucine" OR "(3 amino 2 hydroxy 4 phenylbutyryl) leucine OR bestatine" OR "n (3 amino 2 hydroxy 1 oxo 4 phenylbutyl) leucine" OR "nk 421" OR ubenimex)', yielding 11 results; which were screened by title and abstract. Of these, one (1) article was included in this section.

Results: this study found shows an improvement in resting membrane potentials (RMPs), which almost attained normal levels. Additionally, administration of bestatin significantly reduced the occurrence of electrical myotonia. This article was classified as category I.

Reporting Quality		Risk of Bias	
Parameter (N = 1)	Y (%)	Parameter (N = 1)	Y/U (%)
Type of Facility	0.0	Allocation Concealment	0/100
Type of Cage or Housing	0.0	Blinded Outcome Assessment	0/100
Bedding Material	0.0	Blinded Operations	0/100
N Cage Companions	0.0	Random Cage Allocation	0/100
Breeding Programme	0.0	Random Outcome Assessment	0/100
Light/Dark Cycle	0.0	Sequence Generation	0/100
Temperature and Humidity	0.0	Baseline Characteristics	0/100
Quality of the Water (fish)	-	Incomplete Outcome Data	0/100
Type of Food	0.0	Selective Outcome Reporting	100/0
Access to Food and Water	0.0	Other	0/100
Environmental Enrichment	0.0		
Any Blinding	0.0		
Any Randomisation	0.0		
Sample Size	0.0		
Sample Size Calculation	0.0		
Acclimatisation	0.0		
Sex Disclosed	0.0		
Male/Female/Both (N = 1)	-/-/-		
Background Control	-		
Background Model	-		

7.2.2 Myostatin inhibitors: Stamulumab [72]

Methodology: studies were searched on 19/09/2016 on PubMed with the string 'mdx AND ("myostatin inhibitor" OR "myostatin inhibitors" OR "inhibition of myostatin" OR "myostatin blockade")', yielding 33 results; and on Embase with the string '("duchenne muscular dystrophic mouse" OR "duchenne muscular dystrophy mouse" OR "mdx mouse" OR "mice, inbred mdx" OR "x chromosome-linked muscular dystrophy mouse" OR "x-linked muscular dystrophic mouse" OR "x-linked muscular dystrophy mouse") AND ("myostatin inhibitor" OR "myostatin inhibitors" OR "inhibition of myostatin" OR "myostatin blockade")', yielding 11 results; which were screened by title and abstract. Of these, one (1) articles were included in this section.

Results: this study reported an increase in body weight, muscle mass, size and absolute muscle strength as assessed by the Rota-rod apparat. Moreover, stamulumab decreased muscle degeneration and serum CK levels. This article was classified as category I.

Reporting Quality		Risk of Bias	
Parameter (N = 1)	Y (%)	Parameter (N = 1)	Y/U (%)
Type of Facility	0.0	Allocation Concealment	0/100
Type of Cage or Housing	0.0	Blinded Outcome Assessment	0/100
Bedding Material	0.0	Blinded Operations	0/100
N Cage Companions	0.0	Random Cage Allocation	0/100
Breeding Programme	0.0	Random Outcome Assessment	0/100
Light/Dark Cycle	0.0	Sequence Generation	0/100
Temperature and Humidity	0.0	Baseline Characteristics	0/100
Quality of the Water (fish)	-	Incomplete Outcome Data	0/100
Type of Food	0.0	Selective Outcome Reporting	100/0
Access to Food and Water	0.0	Other	100/0
Environmental Enrichment	0.0		
Any Blinding	0.0		
Any Randomisation	0.0		
Sample Size	0.0		
Sample Size Calculation	0.0		
Acclimatisation	100		
Sex Disclosed	100		
Male/Female/Both (N = 1)	100/0/0		
Background Control	-		
Background Model	-		

7.3 Have drugs with different mechanisms of action and acting on different pathways been tested in this model? If so, which?	Score
Yes, partially.	4.16
Remarks: All the 5 identified mechanisms of action that have been tested in humans have also been tested in mdx mice.	

8. ENDPOINT VALIDATION	
8.1 Are the endpoints used in preclinical studies the same or translatable to the clinical endpoints?	Score
Yes.	6.25
Remarks: most functional measurements of functional pathology in mice, such as grip strength and treadmill exercise are directly correlated to the functional measurements in humans (6-min walking test, ability to stand up without the use of hands etc.).	
8.2 Are the methods used to assess preclinical endpoints comparable to the ones used to assess related clinical endpoints?	Score
Yes.	6.25
Remarks: See question 8.1.	

MODEL NAME	Golden Retriever Muscular Dystrophy (GRMD) dog
INDICATION	Duchenne Muscular Dystrophy
VALIDATION DATE	19.09.2016
TOTAL SUBSECTIONS	33
TOTAL SCORE	65.42
UNCERTAINTY FACTOR (%)	18.2
VALIDATION LEVEL (%)	Highly Validated (81.8)

HISTORICAL BACKGROUND

The Golden Retriever Muscular Dystrophy (GRMD) dog was first identified in 1958, but it only started to be thoroughly characterised in the 80's [73,74]. The mutation in the dog dystrophin gene appeared spontaneously and it was later characterised by Sharp et al, consisting of a point mutation in exon 6, which causes the skipping of exon 7 further leading to the expression of the truncated and non-functional form of dystrophin [75]. Colonies are maintained mostly in academic centres the USA, France, Brazil and Australia [6].

A substantial variation in disease severity, such as the lifespan and ambulation ability of different colonies has been reported in the literature [76]. This should be considered when interpreting the data hereby presented.

1. EPIDEMIOLOGICAL VALIDATION

1.1 Is the model able to simulate the disease in the relevant sexes?	Score
Yes, completely **Remarks**: most literature published on GRMD dogs reports the use of male animals. DMD occurs almost exclusively in males with an incidence of 1 every 3,500 to 6,291 births [9,10]	6.25

1.2 Is the model able to simulate the disease in the relevant age groups (juvenile, adult or ageing)?	Score
Yes, completely **Remarks**: DMD is a disease caused by a genetic defect in the dystrophin coding gene. Like affected boys, GRMD dogs are born with the disease and manifest it in their juvenile phase, often dying in early adulthood [9,77–80].	6.25

2. SYMPTOMATOLOGY AND NATURAL HISTORY (SNH) VALIDATION

2.1 Is the model able to replicate the symptoms and co-morbidities commonly present in this disease? If so, which ones?	Score
Yes, partially.	2

2.1.1 Symptoms modelled

Reduced grown-up body weight: at birth, no difference is seen regarding body weight. Starting at 1 month and further at 6 months, it is possible to notice a reduced body weight when compared to controls of up to 40% [77].

Reduced lifespan: affected dogs live up to ~3 years, a 75% reduction, similar to humans

Muscle wasting: comparable to humans for most muscles with significant fibre necrosis, fibrosis and progressive degeneration [77,78,81–83].

ECG abnormality: characteristic ECG changes include increased Q/R ratios, decreased PR intervals, and frequent ventricular arrhythmias [79].

Cardiomyopathy: myocardial degeneration, fibrosis and mineralisation alongside with decreased cardiac function and congestive heart failure have been reported in GRMD dogs [77,84,85].	

2.1.2 Symptoms partially modelled	

Loss of ambulation: complete loss of ambulation occurs in about 1/3 of GRMD dogs, although dogs which maintain ambulation lose mobility considerably. The reason for this is probably less related to the disease, but more to the fact that dogs are quadrupeds as was suggested that the main reason for the loss of ambulation in humans is more due to the difficulty in balancing the upper body than lower limbs muscle weakness [77,80,86].	

2.1.3 Symptoms not modelled	Score
Cognitive and CNS defects: not reported (unclear).	

2.2 Is the natural history of the disease similar to humans regarding:	

2.2.1 Time to onset;	Score
Yes, completely. **Remarks**: although GRMD dogs have a somewhat high rate of neonatal death (~25%) not seen in either mdx mice or humans [77,78], the first clinical signs appear at 6-9 weeks, analogous to humans at 2-4 years [9,79,80].	2.50

2.2.2 Disease progression;	Score
Yes, partially. **Remarks**: in humans, first symptoms start around 2.5 years, with loss of ambulation around 10 to 15 years [11–13]. In GRMD dogs, muscular atrophy along with joint involvement (including jaw mobility), abnormal gait, decreased respiratory function and cardiomyopathy appears at around 6 months of age [9,77,83]. Similar to DMD, there is a honeymoon phase between 6 and 10 months wherein the disease is relatively stable [9,77,87]. Loss of ambulation is infrequent, with only around 1/3 of GRMD dogs losing ambulation completely, although they do lose mobility considerably [77,80,86].	1.25

2.2.3 Duration of symptoms;	Score
Yes, completely. **Remarks**: symptoms in DMD are lifelong due to permanent muscle degeneration. In GRMD dogs, muscle wasting and degeneration are progressive, leading to permanent damage and significantly reduced lifespan like in humans (9).	2.5

2.2.4 Severity.	Score
Yes, partially. **Remarks**: muscle regeneration in dogs is less prominent than in mice and on par with humans, leading to significant muscle degeneration. The overall loss of muscle strength and function is strikingly similar to humans, however, the infrequency of loss of ambulation makes the manifestation in dogs somewhat less severe than in humans.	1.25

3. GENETIC VALIDATION	
3.1 Does this species also have orthologous genes and/or proteins involved in the human disease?	Score
Yes. **Remarks**: DMD is caused by the lack of dystrophin, thus the dystrophin gene alongside with utrophin (a functional and structural analogue) have been included in this section [31]. The dog has orthologous genes for both of them [88,89].	4.17
3.1.1 Dystrophin	
The human dystrophin gene (Gene ID: 1756) is the largest gene found in humans measuring a total of 2.4 Mb located at Xp21.2-p21.1 [34]. Dystrophin forms parts of the dystrophin-glycoprotein complex (DGC), responsible for connecting the inner cytoskeleton and the extracellular matrix [34]. In dogs, the dystrophin gene (Gene ID: 606758) is located at chromosome X [88].	
3.1.2 Utrophin	
The human utrophin gene (Gene ID: 7402) is located in humans at 6q24.2 [35]. Utrophin is a structural and functional analogue to dystrophin, being also present at neuromuscular synapses and myotendinous junctions [35]. Utrophin upregulation is thought to be a compensation mechanism partially responsible for the milder phenotype in mdx mice [35]. In dogs, the utrophin gene (Gene ID: 442965) is located at chromosome 1 [89].	
3.2 If so, are the relevant genetic mutations or alterations also present in the orthologous genes/proteins?	Score
Yes, partially. **Remarks**: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X [31]. In that sense, DMD being an X-linked inherited disorder, the GRMD dog has the same aetiology (spontaneous mutations) [90]. Nevertheless, GRMD dogs have a point mutation in exon 6, which causes the skipping of exon 7 leading to the truncated form of dystrophin [75]. This is not the case for human patients who have other alterations that lead to dystrophin deficiency.	2.09
3.3 If so, is the expression of such orthologous genes and/or proteins similar to the human condition?	Score
Yes, completely.	4.16
3.3.1 Dystrophin	
Yes, completely. **Remarks**: Like DMD patients, GRMD dogs are dystrophin-deficient not expressing full-length dystrophin [75].	
3.3.2 Utrophin	
Yes, completely. **Remarks**: Utrophin upregulation was reported at the sarcolemma of dystrophic pups and adults, like in humans [36,37,91].	

4. BIOCHEMICAL VALIDATION	
4.1 If there are known pharmacodynamic (PD) biomarkers related to the pathophysiology of the disease, are they also present in the model?	Score
Yes, completely. **Remarks**: high levels of creatine kinase (CK) – a marker of muscle damage, with the muscle-type variants being predominant in plasma are also present in GRMD dogs [38,82,92–94]. Pyruvate Kinase (PK) is also used as a marker of muscle damage and it is present at high levels in DMD boys, however, it has secondary importance (and it is considerably less reliable as a biomarker than CK) as a tracker of disease progression [38]. It is included here for information purposes. It is not clear whether GRMD dogs have high levels of PK.	3.13
4.2 Do these PD biomarkers behave similarly to humans'?	Score
Yes, completely. **Remarks**: CK levels are increased in GRMD dogs, like in humans [38,82,92–94]. There is no information on PK levels in the literature searched.	3.13
4.3 If there are known prognostic biomarkers related to the pathophysiology of the disease, are they also present in the model?	Score
N/A. **Remarks**: Currently, DMD has no prognostic biomarker which is validated and widely used in clinical practice. Many biomarkers are still under exploratory phase. A list of biomarkers being explored can be found in the supplementary material from Guiraud et al, 2015 and in the paper from Hathout et al, 2016 [39,40]	-
4.4 Do these prognostic biomarkers behave similarly to humans'?	Score
N/A.	

5. AETIOLOGICAL VALIDATION	
5.1 Is the aetiology of the disease similar to humans'?	Score
Yes, partially. **Remarks**: DMD is caused by deletions, duplications, small mutations or other smaller rearrangements of the gene that codes for dystrophin in chromosome X [31]. In that sense, DMD being an X-linked inherited disorder, the GRMD dog has the same aetiology (spontaneous mutations) [90]. Nevertheless, GRMD dogs have a point mutation in exon 6, which causes the skipping of exon 7 leading to the truncated form of dystrophin [75]. This is not the case for human patients who have other alterations that lead to dystrophin deficiency.	6.25

6. HISTOLOGICAL VALIDATION	
6.1 Do the histopathological structures in relevant tissues resemble the ones found in humans?	Score
Yes, partially.	8.33
6.1.1 Histopathological features modelled	
Muscle regeneration: existent but in similar intensity as humans, being unable to outpace the increasing muscle necrosis [78,81,92,95,96].	

6.1.2 Histopathological features partially modelled
Limb muscle fibrosis: progressive as seen in the human disease, but tend to plateau in adult life [77,78,81,82,84,92,95,96]
Adipose tissue: fatty infiltration of muscle is less profound in GRMD dogs than in man, sometimes not being evident [77,79,81].

7. PHARMACOLOGICAL VALIDATION	
7.1 Are effective drugs in humans also effective in this model?	Score
	0.58

7.1.1 Glucocorticosteroids

7.1.1.1 Prednisone [55,97,98]

Methodology: studies were searched on 15/09/2016 on PubMed with the string '(GRMD OR "golden retriever muscular dystrophy") AND (Prednisone[mesh] OR prednisone OR Apo-Prednisone OR Cortan OR Cortancyl OR Cutason OR Dacortin OR Decortin OR Decortisyl OR Dehydrocortisone OR Deltasone OR Encorton OR Encortone OR Enkortolon OR Kortancyl OR "Liquid Pred" OR Meticorten OR Orasone OR Panafcort OR Panasol OR Predni Tablinen OR Prednidib OR Predniment OR Prednison Acsis OR "Prednison Galen" OR "Prednison Hexal" OR Pronisone OR Rectodelt OR Sone OR Sterapred OR Ultracorten OR Winpred OR delta-Cortisone)', yielding 4 results; and on Embase with the string '(GRMD OR "golden retriever muscular dystrophy") AND (prednisone OR "1, 2 dehydrocortisone" OR "17, 21 dihydroxypregna 1, 4 diene 3, 11, 20 trione" OR ancortone OR apo-prednisone OR biocortone OR colisone OR cortan OR cortidelt OR cortiprex OR cutason OR dacorten OR "de cortisyl" OR decortancyl OR decortin OR "decortin e merck" OR decortine OR decortisyl OR dehydrocortisone OR dekortin OR delitisone OR "dellacort a" OR "delta 1 dehydrocortisone" OR "delta cortelan" OR "delta cortisone" OR "delta dome" OR "delta e" OR "delta prenovis" OR delta-dome OR deltacorten OR deltacortene OR deltacortisone OR deltacortone OR deltasone OR deltison OR deltisona OR deltra OR "di adreson" OR di-adreson OR diadreson OR drazone OR encorton OR encortone OR enkorton OR fernisone OR hostacortin OR insone OR "liquid pred" OR lodotra OR me-korti OR meprison OR metacortandracin OR meticorten OR meticortine OR nisona OR "nsc 10023" OR nsc10023 OR orasone OR orisane OR panafcort OR paracort OR pehacort OR precort OR precortal OR prednicen-m OR prednicorm OR prednicot OR prednidib OR prednison OR "prednisone alcohol" OR "prednisone intensol" OR "prednisone test" OR prednitone OR "pregna 1, 4 diene 3, 11, 20 trione 17, 21 diol" OR pronison OR pronisone OR pronizone OR pulmison OR rayos OR rectodelt OR servisone OR steerometz OR sterapred OR "sterapred ds" OR ultracorten OR urtilone OR winpred), yielding 6 results; which were screened by title and abstract. Of these, 3 articles were included in this section.

Results: the only study which investigated functional outcomes showed mixed results. High-dose prednisone increased fibre calcification and necrosis, and maximal isomeric tibiotarsal extension strength while paradoxically decreasing maximal isomeric flexion strength. No statistically significant difference between treatment groups and untreated GRMD dogs was found regarding joint contracture, muscle hypertrophy. In the other studies, prednisone was shown to increase α7-integrin, laminin- α2 and γ_{cyto} actin expression. Two (2) articles were classified as category I and one (1) as category II. The latter is partially in line with clinical findings.

Reporting Quality		Risk of Bias	
Parameter (N = 3)	Y (%)	Parameter (N = 3)	Y/U (%)
Type of Facility	0.0	Allocation Concealment	0/100
Type of Cage or Housing	0.0	Blinded Outcome Assessment	0/100
Bedding Material	0.0	Blinded Operations	0/100
N Cage Companions	0.0	Random Cage Allocation	0/100
Breeding Programme	100	Random Outcome Assessment	0/100
Light/Dark Cycle	0.0	Sequence Generation	0/100
Temperature and Humidity	0.0	Baseline Characteristics	0/100
Quality of the Water (fish)	-	Incomplete Outcome Data	0/100
Type of Food	0.0	Selective Outcome Reporting	100/0
Access to Food and Water	0.0	Other	33.3/0
Environmental Enrichment	0.0		
Any Blinding	33.3		
Any Randomisation	0.0		
Sample Size	100		
Sample Size Calculation	33.3		
Acclimatisation	0.0		
Sex Disclosed	33.3		
Male/Female/Both (N = 1)	0/0/100		
Background Control	-		
Background Model	-		

7.1.1.2 Deflazacort

Methodology: studies were searched on 15/09/2016 on PubMed with the string 'GRMD OR "golden retriever muscular dystrophy" AND deflazacort', yielding no results; and on Embase with the string '(GRMD OR "golden retriever muscular dystrophy") AND (deflazacort OR "9 defluorofluazacort" OR azacort OR calcort OR deflan OR defluorofluazacort OR "dl 458" OR "dl 458 it" OR "dl 458it" OR "dl 5458" OR dl458 OR dl458it OR dl5458 OR emflaza OR flantadin OR "fluazacort, difluoro" OR rosilar), yielding 1 result; which was screened by title and abstract. This article was not included in this section.

7.1.2 Read-through compound: Ataluren

Methodology: studies were searched on 15/09/2016 on PubMed with the string 'searched on PubMed with the string 'GRMD OR "golden retriever muscular dystrophy" AND (ptc124 OR ataluren)', yielding no results; and on Embase with the string '(GRMD OR "golden retriever muscular dystrophy") AND ("3 [5 (2 fluorophenyl) 1, 2, 4 oxadiazol 3 yl] benzoic acid" OR "3 [5 (2 fluorophenyl) [1, 2, 4] oxadiazol 3 yl] benzoic acid" OR "ataluren sodium" OR "ptc 124" OR ptc124 OR translarna), yielding 1 result; which was screened by title and abstract. This article was not included in this section.

7.1.3 Exon skipping: Eteplirsen

Methodology: studies were searched on 19/09/2016 on PubMed with the string 'GRMD OR "golden retriever muscular dystrophy" AND '(eteplirsen OR avi-4658 OR "avi 4658" OR AVI-4225 OR "AVI 4225")", yielding no results; and on Embase with the string '(GRMD OR "golden retriever muscular dystrophy") AND ("3 [5 (2 fluorophenyl) 1, 2, 4 oxadiazol 3 yl] benzoic acid" OR "3 [5 (2 fluorophenyl) [1, 2, 4] oxadiazol 3 yl] benzoic acid" OR "ataluren sodium" OR "ptc 124" OR ptc124 OR translarna), yielding no results.

7.2 Are ineffective drugs in humans also ineffective in this model?	Score
Two identified drug classes comprise drugs that were tested in humans and failed due to the lack of efficacy: aminopeptidase and leucine peptidase inhibitors; and myostatin inhibitors.	0.42

7.2.1 Aminopeptidase and leucine peptidase inhibitors: Bestatin
Methodology: studies were searched on 19/09/2016 on PubMed with the string '("golden retriever muscular dystrophy" OR GRMD) AND (Ubenimex[mesh] OR ubenimex OR "uben 3-amino-2-hydroxy-4-phenylbutyryl-L-leucine" OR bestatin OR "bestatin, (D-Leu)-(R-(R*,R*))-isomer" OR "bestatin, (D-Leu)-(R-(R*,S*))-isomer" OR "bestatin, (D-Leu)-(S-(R*,S*))-isomer" OR "bestatin, (L-Leu)-(R-(R*,R*))-isomer OR bestatin, (L-Leu)-(R-(R*,S*))-isomer" OR "bestatin, (L-Leu)-(S-(R*,R*))-isomer" OR "bestatin, (L-Leu)-(S-(R*,S*))-isomer, hydrochloride")', yielding no results; and on Embase with the string '(GRMD OR "golden retriever muscular dystrophy") AND (bestatin OR "(3 amino 2 hydroxy 4 phenylbutanoyl) leucine" OR "(3 amino 2 hydroxy 4 phenylbutyryl) leucine; bestatine" OR "n (3 amino 2 hydroxy 1 oxo 4 phenylbutyl) leucine" OR "nk 421" OR ubenimex), yielding no results.

7.2.2 Myostatin inhibitors: Stamulumab
Methodology: studies were searched on 19/09/2016 on PubMed with the string '("golden retriever muscular dystrophy" OR GRMD) AND (stamulumab OR "myo 029" OR myo029)', yielding no results; and on Embase with the string '(GRMD OR "golden retriever muscular dystrophy") AND (stamulumab OR "myo 029" OR myo029), yielding no results.

7.3 Have drugs with different mechanisms of action and acting on different pathways been tested in this model? If so, which?	Score
Yes, partially. **Remarks**: only one category II study in GRMD dogs with a glucocorticosteroid – prednisone – has been published so far, evidencing the lack of characterisation of this model regarding the other 4 drug classes identified.	0.83

8. ENDPOINT VALIDATION	
8.1 Are the endpoints used in preclinical studies the same or translatable to the clinical endpoints?	Score
Yes. **Remarks**: the endpoints used in the only preclinical study measured the strength of flexion and extension, which can be considered a proxy of muscle strength.	6.25
8.2 Are the methods used to assess preclinical endpoints comparable to the ones used to assess related clinical endpoints?	Score
No. **Remarks**: the measure of strength for both extension and flexion was taken in anesthetised dogs, in which flexion or extension was caused by electrical stimulation of muscles. DMD boys must run a certain distance or perform certain tasks (e.g. get up the floor without using hands) while being awake and fully aware of the tasks.	0

References

1. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can Animal Models of Disease Reliably Inform Human Studies? PLoS Med [Internet]. 2010 Mar 30;7(3):1–8. Available from: https://doi.org/10.1371/journal.pmed.1000245

2. Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans? BMJ Br Med J. 2004;328(7438):514–7.

3. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ Br Med J [Internet]. 2007;334(7586):197–197. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.39048.407928.BE

4. Bebarta V, Luyten D, Heard K. Emergency medicine animal research: Does use of randomization and blinding affect the results? Acad Emerg Med. 2003;10(6):684–7.

5. Peers IS, South MC, Ceuppens PR, Bright JD, Pilling E. Can you trust your animal study data? Nat Rev Drug Discov [Internet]. 2014;13(7):560. Available from: http://dx.doi.org/10.1038/nrd4090-c1

6. McGreevy JW, Hakim CH, McIntosh MA, Duan D. Animal models of Duchenne muscular dystrophy: from basic mechanisms to gene therapy. Dis Model Mech [Internet]. 2015;8(3):195–213. Available from: http://dmm.biologists.org/content/8/3/195.full.pdf

7. Bulfield G, Siller WG, Wight PA, Moore KJ. X chromosome-linked muscular dystrophy (mdx) in the mouse. Proc Natl Acad Sci U S A [Internet]. 1984;81(4):1189–92. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=344791&tool=pmcentrez&rendertype=abstract

8. Sicinski P, Geng Y, Ryder-Cook A, Barnard E, Darlison M, Barnard P. The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science (80-). 1989;244(4912):1578–80.

9. Emery AEH. Population frequencies of inherited neuromuscular diseases-A world survey. Neuromuscul Disord. 1991;1(1):19–29.

10. MedlinePlus. Duchenne muscular dystrophy [Internet]. [cited 2016 Sep 1]. Available from: https://medlineplus.gov/ency/article/000705.htm

11. Ciafaloni E, Fox DJ, Pandya S, Westfield CP, Puzhankara S, Romitti PA, et al. Delayed Diagnosis in Duchenne Muscular Dystrophy: Data from the Muscular Dystrophy Surveillance, Tracking, and Research Network (MD STARnet). J Pediatr [Internet]. 2009;155(3):380–5. Available from: http://dx.doi.org/10.1016/j.jpeds.2009.02.007

12. Bello L, Morgenroth LP, Gordish-Dressman H, Hoffman EP, Mcdonald CM, Cirak S, et al. DMD genotypes and loss of ambulation in the CINRG Duchenne Natural History Study. Neurology. 2016;87(4):401–9.

13. Eagle M, Baudouin S V., Chandler C, Giddings DR, Bullock R, Bushby K. Survival in Duchenne muscular dystrophy: Improvements in life expectancy since 1967 and the impact of home nocturnal ventilation. Neuromuscul Disord. 2002;12(10):926–9.

14. Bostick B, Yue Y, Long C, Duan D. Prevention of dystrophin-deficient cardiomyopathy in twenty-one-month-old carrier mice by mosaic dystrophin expression or complementary dystrophin/utrophin expression. Circ Res. 2008;102(1):121–30.

15. Bostick B, Yue Y, Long C, Marschalk N, Fine DM, Chen J, et al. Cardiac expression of a mini-dystrophin that normalizes skeletal muscle force only partially restores heart function in aged Mdx mice. Mol Ther. 2009;17(2):253–61.

16. Hakim CH, Grange RW, Duan D. The Passive mechanical properties of the extensor digitorum longus muscle are compromised in 2 to 20-month-old mdx mice. J Appl Physiol [Internet]. 2011;1656–63. Available from: http://jap.physiology.org/content/early/2011/03/10/japplphysiol.01425.2010%5Cnhttp://jap.physiology.org/content/early/2011/03/10/japplphysiol.01425.2010.full.pdf

17. Pastoret C, Sebille A. Mdx Mice Show Progressive Weakness and Muscle Deterioration With Age. J Neurol Sci. 1995;129(2):97–105.

18. Chamberlain JS, Metzger J, Reyes M, Townsend D, Faulkner JA. Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma. Faseb J. 2007;21(9):2195–204.

19. Stedman HH, Sweeney HL, Shrager JB, Maguire HC, Panettieri R a, Petrof B, et al. The mdx mouse diaphragm reproduces the degenerative changes of Duchenne muscular dystrophy. Vol. 352, Nature. 1991. p. 536–9.

20. Vaillend C, Billard JM, Laroche S. Impaired long-term spatial and recognition memory and enhanced CA1 hippocampal LTP in the dystrophin-deficient Dmdmdxmouse. Neurobiol Dis. 2004;17(1):10–20.

21. Xu S, Shi D, Pratt SJP, Zhu W, Marshall A, Lovering RM. Abnormalities in brain structure and biochemistry associated with mdx mice measured by in vivo MRI and high resolution localized1H MRS. Neuromuscul Disord. 2015;25(10):764–72.

22. Remmelink E, Aartsma-Rus A, Smit AB, Verhage M, Loos M, van Putten M. Cognitive flexibility deficits in a mouse model for the absence of full-length dystrophin. Genes, Brain Behav. 2016;15(6):558–67.

23. Karagan NJ, Richman LC, Sorensen JP. Analysis of verbal disability in duchenne muscular dystrophy. J Nerv Ment Dis [Internet]. 1980;168(7):419–23. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0018831758&partnerID=40&md5=f07de7d8ddd92b1f82b6c67291405403

24. Dorman C, Hurley AD, D'Avignon J. Language and learning disorders of older boys with Duchenne muscular dystrophy. DevMedChild Neurol [Internet]. 1988;30(3):316–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3402673

25. Leibowitz D, Dubowitz V. Intellect and Behaviour in Duchenne Muscular Dystrophy. Dev Med Child Neurol. 1981;23(6):577–90.

26. Anderson SW, Routh DK, Ionasescu V V. Serial position memory of boys with Duchenne muscular dystrophy. Dev Med Child Neurol [Internet]. 1988;30(3):328–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3402674

27. Lynch GS, Hinkle RT, Chamberlain JS, Brooks S V, Faulkner J a. Force and power output of fast and slow skeletal muscles from mdx mice 6 – 28 months old. J Physiol. 2001;535(2):591–600.

28. Lefaucheur JP, Pastoret C, Sebille A. Phenotype of dystrophinopathy in old mdx mice. Anat Rec [Internet]. 1995;242(1):70–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7604983

29. Grounds MD, Radley HG, Lynch GS, Nagaraju K, De Luca A. Towards developing standard operating procedures for pre-clinical testing in the mdx mouse model of Duchenne muscular dystrophy. Neurobiol Dis. 2008;31(1):1–19.

30. McGeachie JK, Grounds MD, Partridge TA, Morgan JE. Age-related changes in replication of myogenic cells in mdx mice: quantitative autoradiographic studies. J Neurol Sci [Internet]. 1993;119(2):169–79. Available from: http://ac.els-cdn.com/0022510X9390130Q/1-s2.0-0022510X9390130Q-main.pdf?_tid=f9e0a294-2943-11e4-b2f4-00000aacb361&acdnat=1408633446_d3f843b7f01109ce941c131e64dd2d45

31. Ferlini A, Neri M, Gualandi F. The medical genetics of dystrophinopathies: Molecular genetic diagnosis and its impact on clinical practice. Neuromuscul Disord [Internet]. 2013;23(1):4–14. Available from: http://dx.doi.org/10.1016/j.nmd.2012.09.002

32. PubMed. Dmd dystrophin, muscular dystrophy [Mus musculus (house mouse)] [Internet]. [cited 2016 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/13405

33. PubMed. Utrn utrophin [Mus musculus (house mouse)] [Internet]. [cited 2016 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/22288

34. PubMed. DMD dystrophin [Homo sapiens (human)] [Internet]. [cited 2016 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/1756

35. PubMed. UTRN utrophin [Homo sapiens (human)] [Internet]. [cited 2016 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/7402

36. Behr TM, Fischer P, Mudra H, Theisen K, Spes C, Uberfuhr P, et al. Upregulation of utrophin in the myocardium of a carrier of Duchenne muscular dystrophy. Eur Hear J [Internet]. 1997;18(4):699–700. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9129907

37. Fanin M, Melacini P, Angelini C, Danieli GA. Could utrophin rescue the myocardium of patients with dystrophin gene mutations? J Mol Cell Cardiol. 1999;31(0022–2828):1501–8.

38. Zatz M, Rapaport D, Vainzof M, Passos-Bueno MR, Bortolini ER, Pavanello R de CM, et al. Serum creatine-kinase (CK) and pyruvate-kinase (PK) activities in Duchenne (DMD) as compared with Becker (BMD) muscular dystrophy. J Neurol Sci. 1991;102(2):190–6.

39. Guiraud S, Aartsma-Rus A, Vieira NM, Davies KE, van Ommen G-JB, Kunkel LM. The Pathogenesis and Therapy of Muscular Dystrophies. Annu Rev Genomics Hum Genet [Internet]. 2015;(May):1–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26048046

40. Hathout Y, Seol H, Han MHJ, Zhang A, Brown KJ, Hoffman EP. Clinical utility of serum biomarkers in Duchenne muscular dystrophy. Clin Proteomics [Internet]. 2016;13:9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4820909&tool=pmcentrez&rendertype=abstract

41. Wissing ER, Millay DP, Vuagniaux G, Molkentin JD. Debio-025 is more effective than prednisone in reducing muscular pathology in mdx mice. Neuromuscul Disord [Internet]. 2010;20(11):753–60. Available from: http://dx.doi.org/10.1016/j.nmd.2010.06.016

42. Guerron AD, Rawat R, Sali A, Spurney CF, Pistilli E, Cha HJ, et al. Functional and molecular effects of arginine butyrate and prednisone on muscle and heart in the mdx mouse model of duchenne muscular dystrophy. PLoS One. 2010;5(6):1–12.

43. Mizunoya W, Upadhaya R, Burczynski FJ, Wang G, Anderson JE. Nitric oxide donors improve prednisone effects on muscular dystrophy in the mdx mouse diaphragm. Am J Physiol Cell Physiol [Internet]. 2011;300(5):C1065-77. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21270295

44. Baudy AR, Sali A, Jordan S, Kesari A, Johnston HK, Hoffman EP, et al. Non-invasive optical imaging of muscle pathology in mdx mice using cathepsin caged near-infrared imaging. Mol Imaging Biol. 2011;13(3):462–70.

45. McIntosh L, Granberg KE, Brière KM, Anderson JE. Nuclear magnetic resonance spectroscopy study of muscle growth, mdx dystrophy and glucocorticoid treatments: Correlation with repair. NMR Biomed. 1998;11(1):1–10.

46. Shawn RA, Mantsch HH, Anderson JE. Infrared spectroscopy of dystrophic mdx mouse muscle tissue distinguishes among treatment groups. J Appl Physiol. 1996;81(5):2328–35.

47. Hinkle RT, Lefever FR, Dolan ET, Reichart DL, Dietrich JA, Gropp KE, et al. Corticortophin releasing factor 2 receptor agonist treatment significantly slows disease progression in mdx mice. BMC Med. 2007;5(18):1–11.

48. Skrabek RQ, Anderson JE. Metabolic shifts and myocyte hypertrophy in deflazacort treatment of mdx mouse cardiomyopathy. Muscle and Nerve. 2001;24(2):192–202.

49. Anderson JE, Mcintosh LM, Poettcker R. Deflazacort but not prednisone improves both muscle repair and fiber growth in diaphragm and limb muscle in vivo in the mdx dystrophic mouse. Muscle and Nerve. 1996;19(12):1576–85.

The validation of FIMD: insights on Type 2 Diabetes and Duchenne Muscular Dystrophy

50. Hartel J V, Granchelli J a, Hudecki MS, Pollina CM, Gosselin LE. Impact of prednisone on TGF-beta1 and collagen in diaphragm muscle from mdx mice. Muscle Nerve. 2001;24(March):428–32.

51. Markham BE, Kernodle S, Nemzek J, Wilkinson JE, Sigler R. Chronic dosing with membrane sealant poloxamer 188 NF improves respiratory dysfunction in dystrophic mdx and mdx/utrophin$^{-/-}$ mice. PLoS One. 2015;10(8):1–24.

52. Granchelli JA, Pollina C, Hudecki MS. Pre-clinical screening of drugs using the mdx mouse. Neuromuscul Disord. 2000;10(4–5):235–9.

53. Sali A, Guerron AD, Gordish-Dressman H, Spurney CF, Iantorno M, Hoffman EP, et al. Glucocorticoid-treated mice are an inappropriate positive control for long-term preclinical studies in the mdx mouse. PLoS One. 2012;7(4):1–9.

54. Yoon SH, Chen J, Grynpas MD, Mitchell J. Prophylactic pamidronate partially protects from glucocorticoid-induced bone loss in the mdx mouse model of Duchenne muscular dystrophy. Bone [Internet]. 2016;90:168–80. Available from: http://dx.doi.org/10.1016/j.bone.2016.06.015

55. Wuebbles RD, Sarathy A, Kornegay JN, Burkin DJ. Levels of α7 integrin and laminin-α2 are increased following prednisone treatment in the mdx mouse and GRMD dog models of Duchenne muscular dystrophy. Dis Model Mech [Internet]. 2013;6:1175–84. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3759337&tool=pmcentrez&rendertype=abstract

56. de Carvalho SC, Apolinário LM, Matheus SMM, Santo Neto H, Marques MJ. EPA protects against muscle damage in the mdx mouse model of Duchenne muscular dystrophy by promoting a shift from the M1 to M2 macrophage phenotype. J Neuroimmunol [Internet]. 2013;264(1–2):41–7. Available from: http://dx.doi.org/10.1016/j.jneuroim.2013.09.007

57. Anderson JE, Weber M, Vargas C. Deflazacort increases laminin expression and myogenic repair, and induces early persistent functional gain in mdx mouse muscular dystrophy. In: Cell Transplantation. 2000. p. 551–64.

58. St-Pierre SJG, Chakkalakal J V, Kolodziejczyk SM, Knudson JC, Jasmin BJ, Megeney L a. Glucocorticoid treatment alleviates dystrophic myofiber pathology by activation of the calcineurin/NF-AT pathway. FASEB J. 2004;18(15):1937–9.

59. Archer JD, Vargas CC, Anderson JE. Persistent and improved functional gain in mdx dystrophic mice after treatment with L-arginine and deflazacort. FASEB J [Internet]. 2006;20(6):738–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16464957

60. Marques MJ, Oggiam DS, Barbin ICC, Ferretti R, Santo Neto H. Long-term therapy with deflazacort decreases myocardial fibrosis in mdx mice. Muscle and Nerve. 2009;40(3):466–8.

61. Pereira JA, Marques MJ, Santo Neto H. Co-administration of deflazacort and doxycycline: A potential pharmacotherapy for Duchenne muscular dystrophy. Clin Exp Pharmacol Physiol. 2015;42(7):788–94.

62. Ségalat L, Grisoni K, Archer J, Vargas C, Bertrand A, Anderson JE. CAPON expression in skeletal muscle is regulated by position, repair, NOS activity, and dystrophy. Exp Cell Res. 2005;302(2):170–9.

63. Anderson JE, Vargas C. Correlated NOS-Iμ and myf5 expression by satellite cells in mdx mouse muscle regeneration during NOS manipulation and deflazacort treatment. Neuromuscul Disord. 2003;13(5):388–96.

64. Kayali R, Ku JM, Khitrov G, Jung ME, Prikhodko O, Bertoni C. Read-through compound 13 restores dystrophin expression and improves muscle function in the MDX mouse model for duchenne muscular dystrophy. Hum Mol Genet. 2012;21(18):4007–20.

65. Alter J, Lou F, Rabinowitz A, Yin H, Rosenfeld J, Wilton SD, et al. Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. Nat Med [Internet]. 2006;12(2):175–7. Available from: http://www.nature.com/doifinder/10.1038/nm1345

66. Malerba A, Sharp PS, Graham IR, Arechavala-Gomeza V, Foster K, Muntoni F, et al. Chronic systemic therapy with low-dose morpholino oligomers ameliorates the pathology and normalizes locomotor behavior in mdx mice. Mol Ther [Internet]. 2011;19(2):345–54. Available from: http://dx.doi.org/10.1038/mt.2010.261

67. Sharp PS, Bye-a-Jee H, Wells DJ. Physiological characterization of muscle strength with variable levels of dystrophin restoration in mdx mice following local antisense therapy. Mol Ther [Internet]. 2011;19(1):165–71. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3017444&tool=pmcentrez&rendertype=abstract

68. Wu B, Xiao B, Cloer C, Shaban M, Sali A, Lu P, et al. One-year treatment of morpholino antisense oligomer improves skeletal and cardiac muscle functions in dystrophic mdx mice. Mol Ther [Internet]. 2011;19(3):576–83. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-79952189079&partnerID=tZOtx3y1

69. Sazani P, Ness KPV, Weller DL, Poage D, Nelson K, Shrewsbury ASB. Chemical and mechanistic toxicology evaluation of exon skipping phosphorodiamidate morpholino oligomers in mdx mice. Int J Toxicol. 2011;30(3):322–33.

70. FDA. Eteplirsen - Pharmacology Review [Internet]. [cited 2016 Sep 19]. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2016/206488Orig1s000PharmR.pdf

71. Kishi M, Kurihara T, Hidaka T, Kinoshita M. The stabilizing effect of bestatin on the resting membrane potentials of X-linked muscular dystrophy mice. JpnJPsychiatry Neurol. 1990;44(3):595–600.

72. Bogdanovich S, Krag TOB, Barton ER, Morris LD, Whittemore L-A, Ahima RS, et al. Functional improvement of dystrophic muscle by myostatin blockade. Nature. 2002;420(6914):418–21.

73. Meier H, H. M, Meier. Myopathies in the dog. Cornell Vet [Internet]. 1958;48(3):313–30. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emcl1&NEWS=N&AN=0008055495%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emcl1&AN=0008055495%5Cnhttp://digitaal.uba.uva.nl:9003/uva-linker?sid=OVID:embase&id=pmid:&id=doi:&

74. Valentine BA, Cooper BJ, Cummings JF, deLahunta A. Progressive muscular dystrophy in a golden retriever dog: light microscope and ultrastructural features at 4 and 8 months. Acta Neuropathol [Internet]. 1986;71(3–4):301–10. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3799143

75. Sharp NJH, Kornegay JN, Van Camp SD, Herbstreith MH, Secore SL, Kettle S, et al. An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. Genomics. 1992;13(1):115–21.

76. Kornegay JN. The golden retriever model of Duchenne muscular dystrophy. Skelet Muscle. 2017;7(9):1–21.

77. Valentine BA, Cooper BJ, de Lahunta A, O'Quinn R, Blue JT. Canine X-linked muscular dystrophy. An animal model of Duchenne muscular dystrophy: Clinical studies. J Neurol Sci. 1988;88(1–3):69–81.

78. Valentine BA, Cooper BJ. Canine X-linked muscular dystrophy: Selective involvement of muscles in neonatal dogs. Neuromuscul Disord. 1991;1(1):31–8.

79. Valentine B a, Winand NJ, Pradhan D, Moise NS, de Lahunta a, Kornegay JN, et al. Canine X-linked muscular dystrophy as an animal model of Duchenne muscular dystrophy: a review. Am J Med Genet [Internet]. 1992;42(3):352–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1536178

80. Barthélémy I, Pinto-Mariz F, Yada E, Desquilbet L, Savino W, Silva-Barbosa SD, et al. Predictive markers of clinical outcome in the GRMD dog model of Duchenne muscular dystrophy. Dis Model Mech [Internet]. 2014;7(11):1253–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4213729%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1045041%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/25261568%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4213729

81. Valentine BA, Cooper BJ, Cummings JF, de Lahunta A. Canine X-linked muscular dystrophy: Morphologic lesions. J Neurol Sci. 1990;97(1):1–23.

82. McCully K, Giger U, Argov Z, Valentine B, Cooper B, Chance B, et al. Canine X-linked muscular dystrophy studied with in vivo phosphorus magnetic resonance spectroscopy. Muscle Nerve. 1991;14(11):1091–8.

83. Valentine BA, Cooper BJ, Gallagher EA. Intracellular calcium in canine muscle biopsies. J Comp Pathol. 1989;100(3):223–30.

84. Cooper BJ, Winand NJ, Stedman H, Valentine BA, Hoffman EP, Kunkel LM, et al. The homologue of the Duchenne locus is defective in X-linked muscular dystrophy of dogs. Nature. 1988;334(6178):154–6.

85. Valentine B a, Cummings JF, Cooper BJ. Development of Duchenne-type cardiomyopathy. Morphologic studies in a canine model. Am J Pathol. 1989;135(4):671–8.

86. Barthelemy I, Barrey E, Aguilar P, Uriarte A, Le Chevoir M, Thibaud JL, et al. Longitudinal ambulatory measurements of gait abnormality in dystrophin-deficient dogs. Bmc Musculoskelet Disord. 2011;12:1–11.

87. Fan Z, Wang J, Ahn M, Shiloh-Malawsky Y, Chahin N, Elmore S, et al. Characteristics of magnetic resonance imaging biomarkers in a natural history study of golden retriever muscular dystrophy. Neuromuscul Disord [Internet]. 2014;24(2):178–91. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4065593&tool=pmcentrez&rendertype=abstract

88. PubMed. DMD dystrophin [Canis lupus familiaris (dog)] [Internet]. [cited 2016 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/606758

89. PubMed. UTRN utrophin [Canis lupus familiaris (dog)] [Internet]. [cited 2016 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/gene/442965

90. Cooper BJ, Valentine BA, Wilson S, Patterson DF, Concannon PW. Canine muscular dystrophy: confirmation of X-linked inheritance. J Hered [Internet]. 1988;79(6):405–8. Available from: http://jhered.oxfordjournals.org/content/79/6/405.full.pdf

91. Lanfossi M, Cozzi F, Bugini D, Colombo S, Scarpa P, Morandi L, et al. Development of muscle pathology in canine X-linked muscular dystrophy. I. Delayed postnatal maturation of affected and normal muscle as revealed by myosin isoform analysis and utrophin expression. Acta Neuropathol. 1999;97(2):127–38.

92. Kornegay JN, Tuler SM, Miller DM, Levesque DC. Muscular dystrophy in a litter of golden retriever dogs. Muscle Nerve. 1988;11(10):1056–64.

93. Valentine BA, Blue JT, Cooper BJ. The effect of exercise on canine dystrophic muscle. Ann Neurol [Internet]. 1989 Oct [cited 2016 Sep 16];26(4):588. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2817833

94. Valentine B, Blue J, Shelley S, Cooper B. Increased Serum Alanine Aminotransf erase Activity Associated With Muscle Necrosis in the Dog. J Vet Iternal Med. 1990;4:140–3.

95. Nguyen F, Cherel Y, Guigand L, Goubault-Leroux I, Wyers M. Muscle lesions associated with dystrophin deficiency in neonatal golden retriever puppies. J Comp Pathol. 2002;126(2–3):100–8.

96. Cozzi F, Cerletti M, Luvoni GC, Lombardo R, Brambilla PG, Faverzani S, et al. Development of muscle pathology in canine X-linked muscular dystrophy. II. Quantitative characterization of histopathological progression during postnatal skeletal muscle development. Acta Neuropathol [Internet]. 2001;101(5):469–78. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11484818

97. Liu JMK, Okamura CS, Bogan DJ, Bogan JR, Childers MK, Kornegay JN. Effects of prednisone in canine muscular dystrophy. Muscle and Nerve. 2004;30(6):767–73.

98. Hanft LM, Bogan DJ, Mayer U, Kaufman SJ, Kornegay JN, Ervasti JM. Cytoplasmic γ-actin expression in diverse animal models of muscular dystrophy. Neuromuscul Disord. 2007;17(7):569–74.
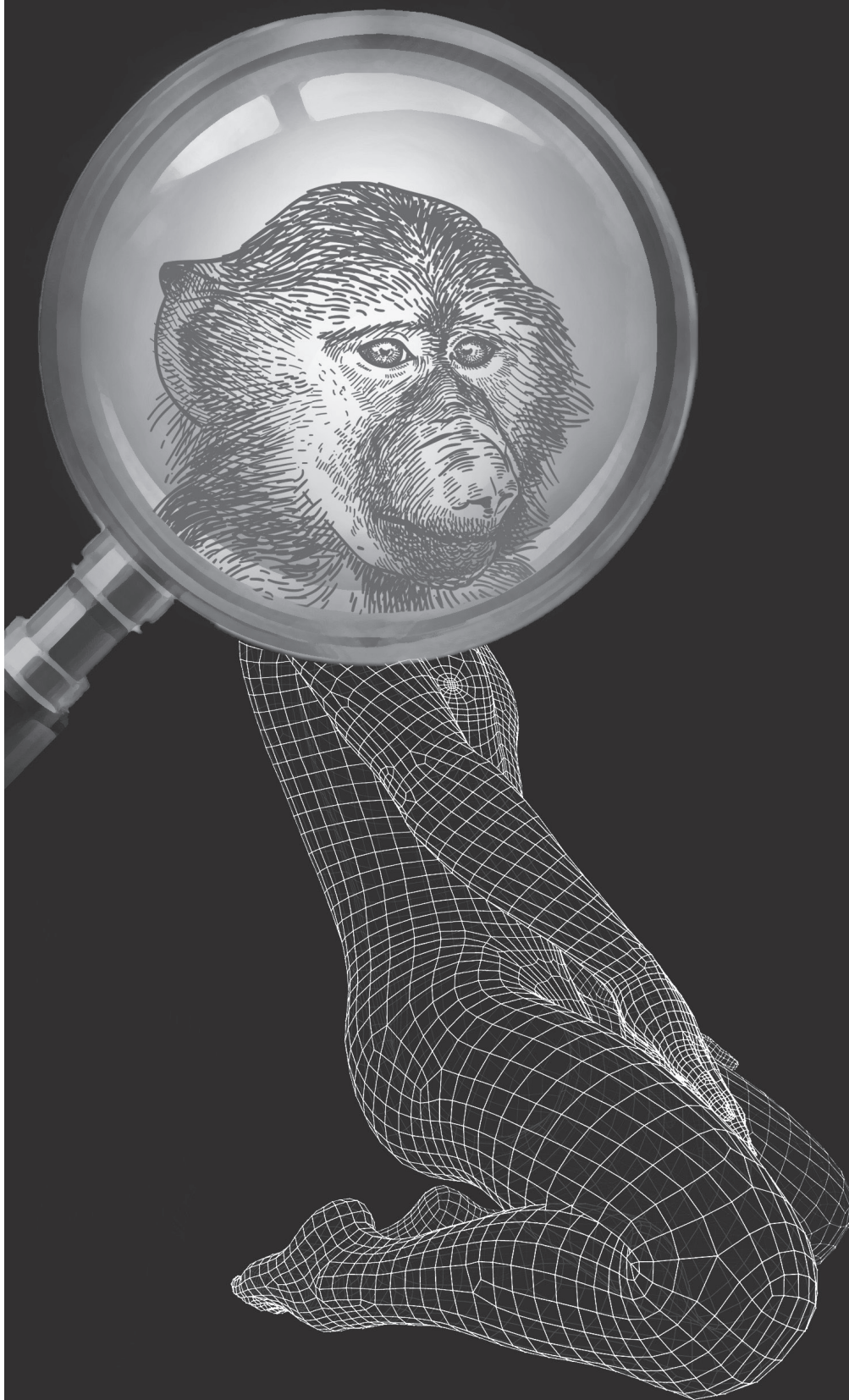
1) Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands;
2) Copernicus Institute of Sustainable Development, Innovation Studies, Utrecht University, Utrecht, The Netherlands;
3) Department for Health Evidence Unit SYRCLE, Radboud University Medical Centre, Nijmegen, The Netherlands;
4) Department of Anesthesiology, Radboud University Medical Centre, Nijmegen, The Netherlands;
5) Medicines Evaluation Board, Utrecht, The Netherlands. `

4

Guilherme S. Ferreira[1],
Désirée H. Veening-Griffioen[1],
Wouter P.C. Boon[2],
Carlijn R. Hooijmans[3,4],
Ellen H.M. Moors[2],
Huub Schellekens[1],
Peter J.K. van Meer[1,5]
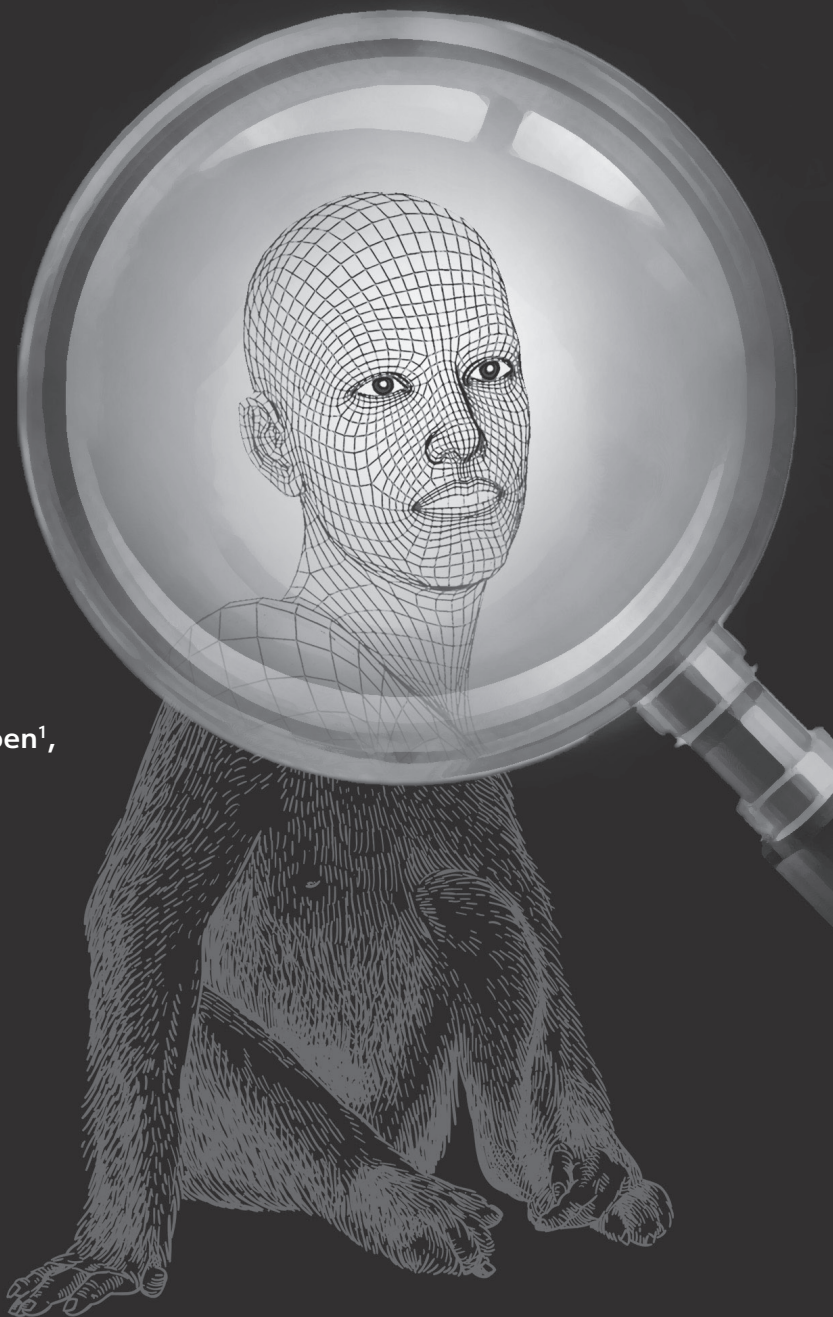
COMPARISON OF DRUG EFFICACY IN TWO ANIMAL MODELS OF TYPE 2 DIABETES: A SYSTEMATIC REVIEW AND META-ANALYSIS

Abstract

Previous qualitative research has suggested there are only minor differences between the db/db mouse and the Zucker Diabetic Fatty (ZDF) rat, both animal models of type 2 diabetes. However, it is not known whether these models are also comparable regarding drug response in quantitative terms (effect size). To investigate the extent of these differences, we conducted a systematic review and meta-analysis of approved drugs in these models. We searched on PubMed and Embase on 3.7.2019 for studies including either model, a monotherapy arm with an EMA/FDA approved drug for the treatment of type 2 diabetes, HbA1c assessment and a control group. Studies aimed at diabetes prevention or with surgical interventions were excluded. We calculated the Standardised Mean Difference (SMD) to compare effect sizes (HbA1c reduction) per drug and drug class across models. We included a risk of bias assessment for all included publications. A total of 121 publications met our inclusion criteria. For drugs with more than two comparisons, both models predicted the direction of the effect regarding HbA1c levels. There were no differences between the db/db mouse and ZDF rat, except for exenatide (P = 0.02) and GLP-1 agonists (P = 0.03) in which a larger effect size was calculated in the ZDF rat. Our results indicate the differences between the db/db mouse and ZDF rat are not relevant for preliminary efficacy testing. This methodology can be used to further differentiate between animal models used for the same indication, facilitating the selection of models more likely to predict human response.

Keywords: animal model; drug development; type 2 diabetes; translational research; systematic review; meta-analysis

Introduction

The high attrition rate in drug development has been a topic of debate for over a decade (1). One of the reasons suggested for such low odds to reach the market is the poor translation of animal to clinical data (1,2). Efficacy is the leading cause of failure for more than half of all drugs in development, followed by commercial reasons and safety (3,4).

For type 2 diabetes, the success rate is particularly low: 9.3% – lower than neurology (9.4%) and the average of all indications (10.4%) (3). Thirty-three (33) drugs have been approved so far by the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) to treat type 2 diabetes. Nevertheless, given the prevalence and socioeconomic cost of type 2 diabetes, which affects 32.7 million people in EU and is projected to affect 629 million worldwide by 2045, the quest for new treatments is far from completed (5,6).

To address the issue of poor translational of preclinical data, we previously developed the Framework to Identify Models of Disease (FIMD), a question-based approach to assess, compare and validate animal models of disease (7,8). In FIMD's pilot study, we included two of the most commonly used models in type 2 diabetes drug development: the db/db mouse and the Zucker Diabetic Fatty (ZDF) rat. While both present a mutation in the leptin receptor gene, the ZDF rat also has a defect in β-cell transcription that contributes to the diabetic phenotype (9).

The preliminary validation in FIMD's pilot study showed that, qualitatively, there are only minor differences between the two type 2 diabetes models: they scored virtually the same in all but one domain (Epidemiology, Symptomatology and Natural History, Biochemistry, Aetiology, Histology, Pharmacology and Endpoints). The only exception was the Genetics domain, in which the ZDF rat scored higher than the db/db mouse not because it mimics this aspect of type 2 diabetes more closely, but because there were not as much data available in the literature for the db/db mouse. These results are corroborated by a high similarity factor of almost 90%, a measure of how often questions got the same response in both models.

In this article, we go a step further and conducted a systematic review and meta-analysis to crosscheck the extent of these (dis)similarities in a quantitative fashion. By comparing the effect sizes of different glucose-lowering approved drugs and drug classes on each model, we can provide additional validation for FIMD's ability to differentiate between models in the same indication. We selected the glycosylated haemoglobin A1c (HbA1c) as our primary outcome because it correlates with blood glucose from the previous eight to twelve weeks, and it is the most reliable endpoint

for the assessment of the efficacy of new glucose-lowering drugs according to EMA and FDA (10–13). Additionally, this systematic review can substantiate the basis for a more quantitative approach to discriminate between animal models in any therapeutic area.

Material and Methods

We conducted this systematic review and meta-analysis according to the protocol registered in advance on PROSPERO (https://www.crd.york.ac.uk/PROSPERO/) with ID CRD42019141896.

Search Strategy and Paper Selection

We searched PubMed and Embase on 3.7.2019 for studies investigating the effect of glucose-lowering therapy on HbA1c in db/db mice and ZDF rats. We designed a comprehensive search strategy using three search components (model, indication and drugs), which are detailed in Supporting Information S1[8]. There were no language nor date restrictions. We included studies 1) conducted in either the db/db mouse (C57BLKS/J background or undisclosed) or ZDF rat; 2) that included at least one drug approved up to 3.7.2019 by EMA or FDA to treat type 2 diabetes as monotherapy; 3) using a placebo control; 4) that reported the effect of the monotherapy on HbA1c. We excluded reviews, *in silico*, *in vitro*, *ex vivo*, and clinical studies; and papers on diabetes prevention or with any surgical procedure. We excluded papers that did not report the db/db mouse's background strain but used C57BL/6J littermates or controls, since this background is associated with only mild diabetic symptoms, such as transient hyperglycaemia (14). GSF and DVG independently screened articles by title and abstract using Rayyan — a web and mobile app for systematic reviews (15). The selected papers were then read in full-text, and studies which met all inclusion but no exclusion criteria were included. Any divergences were solved by consulting CH as a third reviewer. We removed duplicates with Rayyan.

Study Characteristics and Data Extraction

GSF and DVG independently extracted the first author, publication year, animal model, background strain, sex, age at the start of treatment, age at the end of treatment, treatment duration, intervention(s), formulation (if available), route of administration, dose, baseline HbA1c (if available), endpoint HbA1c, standard deviation (S.D.) or standard error of the mean (S.E.M.), and number of animals in each study arm. The age at the start of the treatment was calculated by adding the

---

[8]   Available online at https://doi.org/10.1016/j.ejphar.2020.173153.

acclimatisation period to the age at the time of arrival. The age at the end of the treatment was calculated by adding the treatment duration to the age at the start of the treatment. When drugs were dosed less often than daily, we calculated the daily dose by dividing the dose by the interval (in days) between doses. If only graphical data were available, we used a digital ruler to extract the numerical values for the relevant data points whenever possible (16). If the S.D./S.E.M. was not visible, we measured the distance from the middle of the datapoint object (e.g. circle, diamond) to its edge to obtain a conservative estimate. If the study characteristics relevant for the planned meta-analysis were missing (endpoint HbA1c, S.D./S.E.M. or the number of animals), we contacted the corresponding authors. After two weeks, if no response was received, we sent a reminder. If the corresponding address was not valid or if the authors did not reply after a month, we excluded the study from the meta-analysis.

Assessment of Methodological Quality and Risk of Bias

GSF and DVG independently performed the quality assessment using SYRCLE's risk of bias tool (17). A 'Yes' (Y) indicates a low risk of bias, while a 'No' (N) indicates a high risk of bias in a specific domain. Whenever there was not enough information to evaluate the risk of a particular bias, we used 'Unclear' (U). In an attempt to get a more nuanced picture, we added two reporting parameters to our risk of bias assessment: blinding and randomisation at any level. For these topics, a 'Yes' (Y) means the authors mentioned either blinding or randomisation.

Data Synthesis and Statistical Analysis

Data were analysed using the Comprehensive Meta-Analysis software (CMA version 2.0). We converted S.E.M. to S.D. using the formula below for both the control and treatment groups:

$$\text{S.D.} = S.E.M. \times \sqrt{n}$$

If the same control group served more than one treatment group, we divided the number of control animals by the total number of groups. If the number of animals was reported as a range (e.g. 8-11), we used the lowest number (8). We calculated the Standardised Mean Difference (SMD) per model for each drug. The SMD is the result of the difference between the mean from the experimental and control groups divided by their pooled standard deviation. Despite the anticipated heterogeneity, the individual drug effect sizes were pooled to obtain an overall SMD and 95% confidence interval for all drug classes for which HbA1c levels were available for more than one drug. We used the random-effects model, which considers the precision of individual studies and the variation between studies, weighting each study accordingly.

Subgroup analyses were pre-defined in the protocol and performed to assess the influence of variables on the effect size. The results from subgroup analyses were only interpreted when subgroups contained at least three independent studies or five comparisons. We planned subgroup analysis for age at the start and end of treatment, treatment duration, route of administration, dose, and risk of bias. We performed the meta-analysis for each drug and drug class per model as a subgroup analysis. We calculated the extent of heterogeneity using I^2, which describes the variance that we can assign to the differences between studies (18).

Publication Bias

We assessed the publication bias via visual assessment of a funnel plot, Duval and Tweedie's 'trim and fill' analysis, and Egger's regression analysis for drugs with more than ten studies per model. Because SMDs may cause funnel plot distortion, we plotted the SMD against a sample size-based precision estimate:

$$\frac{1}{\sqrt{n}}$$

Whenever we performed multiple comparisons, we used the Holm-Bonferroni method to correct for it.

Sensitivity Analysis

Although subgroup analyses are only exploratory, we investigated the differences between subgroups using sensitivity analysis. We excluded the studies with the characteristic(s) thought to explain these differences to verify whether the results were robust.

Results

We made a few divergences from the pre-registered protocol. Initially, we planned to calculate the Weighted Mean Difference (WMD) because we expected all articles to report HbA1c in the percentage of total haemoglobin. However, some studies reported the HbA1c in other units (e.g. ng/ml or mmol/L). Also, calculating the SMD instead of the WMD allows for a better interpretation given the high heterogeneity usual for animal research.

We expected to conduct subgroup analysis of the route of administration, treatment duration, age at the start and end of treatment, dose and risk of bias. Most drugs were administered via the same route, and therefore, a subgroup analysis was no longer warranted. Since the age at the end of the treatment was dependent on

the age when animals started receiving the intervention and treatment duration, it would not add any additional information. We also excluded the dose from subgroup analysis as it was highly variable in terms of the order of magnitude for the drugs that had enough references/comparisons to be eligible. We could not assess the risk of bias for the vast majority of studies. Thus, we opted not to perform a subgroup analysis of any of the risk of bias parameters. At first, we also did not expect a sensitivity analysis would be necessary. Nonetheless, since we found significant differences between subgroups, we conducted a sensitivity analysis in an attempt to explain these differences.

Study Selection Process

The search on PubMed and Embase retrieved 3,427 and 4,914 publications, respectively (8,341 total). After removal of duplicates, 5,405 abstracts were assessed (35.2% duplicates), and 1,073 were selected for full-text assessment. A total of 121 publications (163 comparisons) met our inclusion criteria. The PRISMA flowchart is shown in Fig 1.
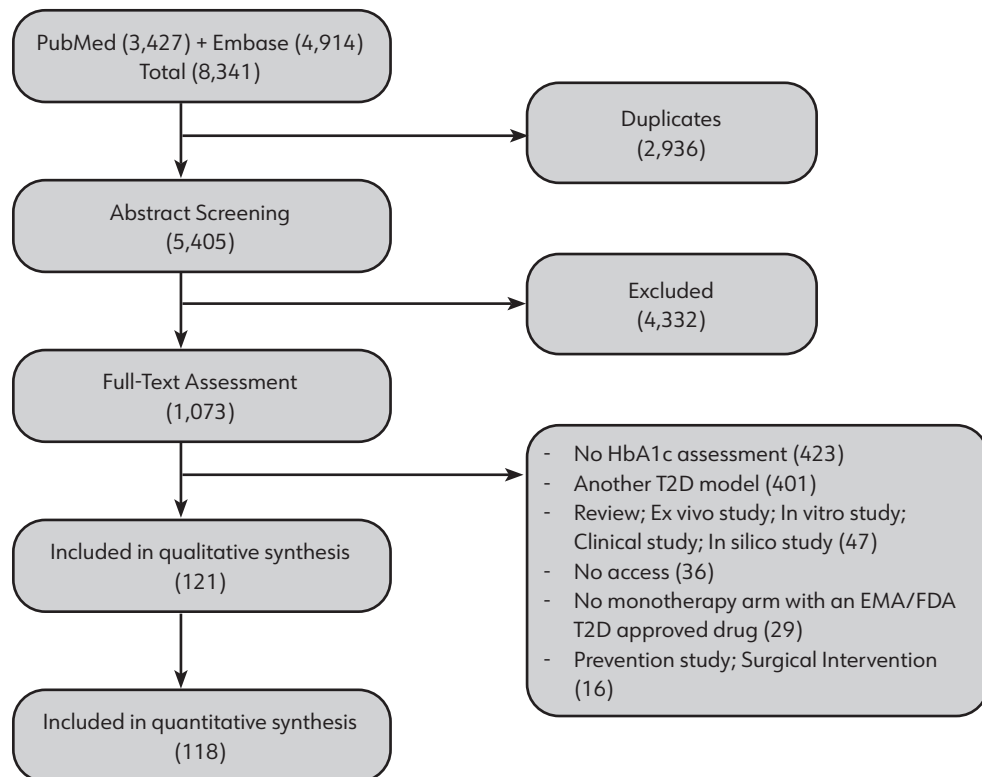

FIGURE 1 | Flowchart of the study selection process.

Study Characteristics

We provide a table with the individual study characteristics in Supporting Information S2[9]. The majority of the included publications used the db/db mouse (n=69) while the ZDF rat was used in 52 publications. Authors mentioned the background of the db/db mouse in 71% of articles (n=69). We retrieved at least one study for 17 approved drugs (acarbose, alogliptin, canagliflozin, colesevelam, dapagliflozin, dulaglutide, empagliflozin, exenatide, glibenclamide, glimepiride, linagliptin, liraglutide, metformin, pioglitazone, rosiglitazone, sitagliptin, vildagliptin). The five drugs with the most comparisons (n=163) included across both models were exenatide (34), rosiglitazone (30), liraglutide (21), metformin (16), and pioglitazone (15).

The majority of comparisons used male animals (88.3%, n=163), followed by unspecified sex (7.4%) and females (7.4%). Over half of the experiments (67.5%) were performed in animals during the development of diabetes (defined as 4-8 weeks for the db/db mouse and 6-12 weeks for the ZDF rat (19,20). Since there was one study with rats aged five weeks with a similar magnitude of the effect, we included it in the 'developing diabetes' group. The treatment duration varied from 2 to 28 weeks, with most treatments lasting from 4.1 to 12 weeks (59.5%) followed by the shortest durations <4 weeks (31.9%). The most frequent routes of administration were oral (62%) and subcutaneous (32.5%).

Study Quality and Risk of Bias

The results of the risk of bias evaluation are presented in Fig 2. No studies mentioned blinding at any level, while 43% (n=121) reported randomisation somehow. There was no report of allocation concealment, blinded outcome assessment, blinded operations or random outcome assessment. The vast majority of publication could not be assessed for random cage allocation or sequence generation (98.3% and 97.5%, respectively). Animal groups were controlled for baseline characteristics (e.g. glycaemia) in only 21.5% of studies. More than half of the articles (52.9%) did not report/justify whether the number of animals was the same before and after the experiment was completed. All studies reported HbA1c levels for all groups mentioned in the methods section. In two cases, we identified an additional source of bias: the first reported significantly different starting HbA1c levels across treatment groups and the second did not replicate the same treatment procedure in the control and experimental groups.

---

[9]   Available online at https://doi.org/10.1016/j.ejphar.2020.173153.
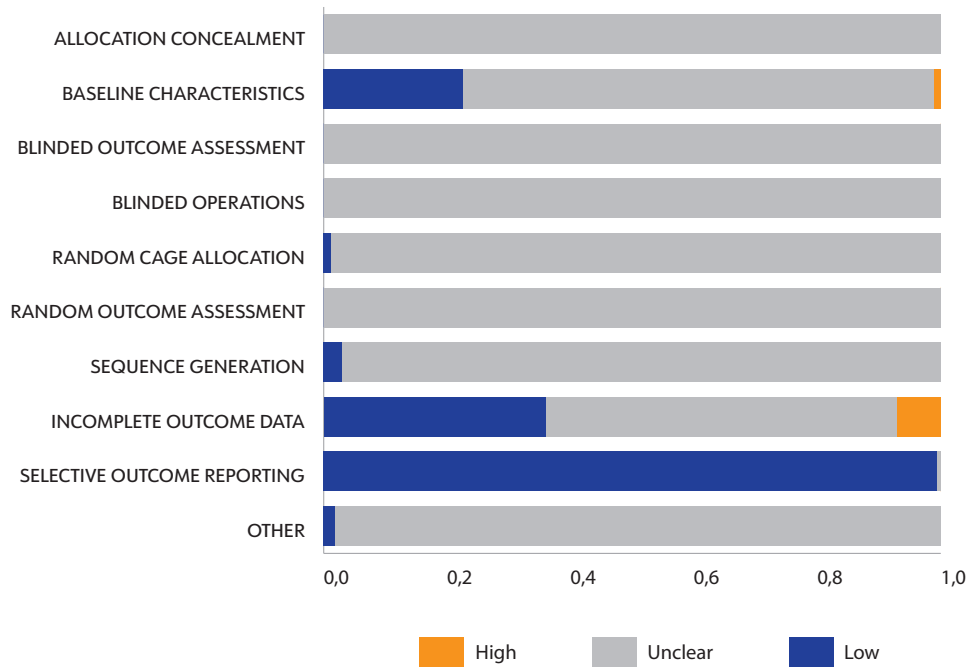
FIGURE 2 | Results of the risk of bias assessment (n=121) according to the SYRCLE Risk of Bias Tool.

Meta-Analysis of Glucose-Lowering Effect on HbA1c

Due to missing data, we excluded three papers (21–23) from the meta-analysis. The remaining 118 studies containing 163 comparisons were included. Table 1 shows the results of the meta-analysis for individual drugs and drug classes (whenever data for two or more drugs were available) for both models.

Empagliflozin, exenatide, liraglutide, metformin, pioglitazone, rosiglitazone and sitagliptin reduced the HbA1c levels in both models. The following drugs did not reach statistical significance but had at most two studies: glimepiride and vildagliptin (db/db mouse); and colesevelam, dapagliflozin, glimepiride, glipizide and vildagliptin (ZDF rat).

In the db/db mouse, the highest SMD was from acarbose (-4.65 [-9.13, -0.16], n=3, I^2 = 94%), followed by rosiglitazone (-3.88 [-4.70, -3.06], n=25, I^2 = 84%). Starting the treatment during the time in which mice were developing diabetes (4-8 weeks, -2.58 [-3.04, -2.12], n=56, I^2 = 84%) had a larger effect size than the ones starting after diabetes was established (older than eight weeks, -1.68 [-2.19, -1.17], n=23, I^2 = 79%) with P = 0.01. A longer treatment duration was also associated with a higher effect size: -2.88 [-3.43, -2.33, n=50, I^2 = 88%] (4.1-12 weeks) when compared to the

shortest duration (0-4 weeks): -1.74 [-2.07, -1.41], n=29, I² = 51%), P = 0.001. There was no difference between sexes (P = 0.09).

TABLE 1 | Standardised Mean Differences (SMDs) per drug, drug class, and animal model.

Drug/Drug Class	Db/db mouse		ZDF rat		Comparison
	SMD (CI 95%)	References (Comparisons)	SMD (CI 95%)	References (Comparisons)	
α-Glucosidase Inhibitors					
Acarbose	-4.65 [-9.13, -0.16]	3 (3)			
Biguanides					
Metformin	-1.44 [-2.36, -0,52]	8 (8)	-2.16 [-2.92, -1,40]	8 (8)	P = 0.25
Bile Acid-Binding Resins					
Colesevelam			1.56 [0.92, 2.20]	1 (1)	
DPP-IV Inhibitors	-0.81 [-1.52, -0.11]	9 (9)	-1.11 [-1.47, -0.75]	9 (9)	P = 0.53
Alogliptin	-0.64 [-1.27, -0.02]	3(3)			
Linagliptin	-1.75 [-3.50, -0.001]	2 (2)			
Sitagliptin	-1.09 [-1.92, -0.27]	2 (2)	-1.36 [-1.79, -0.94]	7 (7)	
Vildagliptin	-0.11 [-2.08, 1.86]	2 (2)	-0.53 [-1.44, 0.37]	2 (2)	
GLP-1 Agonists	-1.49 [-1.94, -1.04]	18 (20)	-2.23 [-2.73, -1.74]	19 (38)	P = 0.03
Dulaglutide			-1.53 [-2.77, -0.28]	1 (3)	
Exenatide	-1.15 [-1.79, -0.52]	9 (11)	-2.32 [-3.01, -1.63]	11 (23)	P = 0.046
Liraglutide	-1.87 [-2.42, -1.32]	9 (9)	-2.42 [-3.35, -1,50]	8 (12)	P = 0.32
SGLT-2 Inhibitors	-2.70 [-3.64, -1.76]	4 (6)	-3.37 [-4.43, -2.32]	7 (12)	P = 0.36
Canagliflozin			-5.59 [-6.91, -4,27]	2 (4)	
Dapagliflozin	-2.87 [-4.43, -1,32]	2 (4)	-3.93 [-8.46, 0.60]	2 (2)	
Empagliflozin	-2.55 [-3.50, -1.59]	2 (2)	-2.23 [-3.37, -1.09]	3 (6)	
Sulphonylureas			-1.03 [-2.81, 0.75]	2 (2)	
Glibenclamide			-0.23 [-1.09, 0.62]	1 (1)	
Glimepiride	-0.94 [-2.13, 0.25]	2 (2)	-2.07 [-3.63, -0.51]	1 (1)	
Thiazolidinediones	-3.63 [-4.29, -2.98]	28 (33)	-3.98 [-5.38, -2.59]	12 (12)	P = 0.74
Pioglitazone	-3.05 [-4.04, -2.06]	8 (8)	-3.10 [-4.69, -1.51]	7 (7)	P = 0.84
Rosiglitazone	-3.88 [-4.70, -3.06]	20 (25)	-5.19 [-7.92, -2.47]	5 (5)	P = 0.37

CI – confidence interval; DPP-IV - Dipeptidyl Peptidase-4; ES – effect size; GLP-1 – Glucagon-Like Peptide-1; SGLT-2 – Sodium/Glucose Cotransporter-2.

In the ZDF rat, canagliflozin had the largest point estimate (-5.59 [-6.91, -4,27], n=4, I^2 = 0%), also followed by rosiglitazone (-5.19 [-7.92, -2.47], n=5, I^2 = 88%). There was no significant difference in effect size for all other characteristics: age at the start of the treatment (P = 0.47), treatment duration (P = 1) or sex (P = 0.72).

The limited number of studies per drug (three independent references or five comparisons) only allowed the comparison of exenatide, liraglutide, metformin, pioglitazone and rosiglitazone between models. Although all five drugs showed a more substantial HbA1c reduction in the ZDF rat compared to the db/db mouse, only for exenatide was this difference statistically significant (P = 0.02).

We calculated pooled SMDs for DPP-IV inhibitors, GLP-1 agonists, SGLT-2 inhibitors, sulphonylureas and thiazolidinediones. In the db/db mouse, all drug classes effectively reduced the HbA1c levels, the thiazolidinediones showed the largest effect (-3.63 [-4.29, -2.98], n=33, I^2 = 81%). In the ZDF rat, sulphonylureas was the only class not to impact HbA1c (-0.40 [-1.13, 0.33], n=4, I^2 = 47%) while the thiazolidinediones also caused the greatest reduction (-3.98 [-5.38, -2.59], n=12, I^2 = 88%). The differences in drug classes between models were negligible, except for GLP-1 agonists with P = 0.03.

Sensitivity Analysis

We explored the differences between the subgroups in the db/db mouse. Since the effect sizes of thiazolidinediones (pioglitazone and rosiglitazone) were significantly higher, and they were more prevalent in the db/db mouse, we investigated whether they were skewing the meta-analysis results. When we removed all studies that used thiazolidinediones, we found no differences in age groups (P = 0.69) nor treatment duration (P = 0.38).

Publication Bias

We assessed the publication bias with a funnel plot and Duval and Tweedie's 'trim and fill'. Only exenatide for the ZDF rat and rosiglitazone for the db/db mouse could be assessed as individual drugs (more than 10 comparisons). Publication bias of GLP-1 agonists and thiazolidinediones were assessed for both models while SGLT-2 inhibitors were assessed only for the ZDF rat. We found no evidence of publication bias (Table 2).

A systematic review and meta-analysis of two animal models of Type 2 Diabetes

TABLE 2 | results from the publication bias funnel plot and 'trim and fill' analysis.

Drug	Model	Egger Regression p-value	Imputed (trim and fill)
Exenatide	ZDF rat	0.27	0
Rosiglitazone	Db/db mouse	0.76	0
Drug Class		**Egger Regression p-value**	**Imputed (trim and fill)**
GLP-1 Agonists	Db/db mouse	0.85	0
GLP-1 Agonists	ZDF rat	0.06	0
SGLT-2 Inhibitors	ZDF rat	0.29	0
Thiazolidinediones	Db/db mouse	0.74	0
Thiazolidinediones	ZDF rat	0.94	1

GLP-1 – Glucagon-Like Peptide-1; SGLT-2 – Sodium/Glucose Cotransporter-2.

Discussion

There are over 30 animal models for type 2 diabetes reported in the literature, and selecting which is the adequate choice to test the preliminary efficacy of new drugs can be challenging (24). Although conducting a systematic review and meta-analysis of animal studies is not new, this is the first report of the use of this methodology to compare diabetes models' ability to predict human efficacy. The systematic reviews and data syntheses published so far focus on investigating a specific intervention or occurrence of adverse events rather than on the models themselves (e.g. (25–27)). The most advanced approach so far was published by Varga and colleagues who presented an innovative method to estimate the predictive validity of animal experiments using rosiglitazone (28). However, their approach covered a small set of publications and did not investigate the risk of bias nor differences between methodologies used in animal and clinical studies. As such, while this method has several merits, it does not consist of a robust way to compare animal models.

FIMD's pilot study research indicated there are only minor differences between the db/db mouse and the ZDF rat, which was expected given the similarity of the aetiologies in both models (7). FIMD assesses models based on eight domains: epidemiology, symptomatology and natural history, genetics, biochemistry, aetiology, histology, pharmacology and endpoints. In two of these domains, genetics (related to genes, genetic alterations and expression) and pharmacology (related to the response to effective and ineffective drugs), the scarcity of studies prevented any further statement as to whether these models are potentially equivalent.

In this systematic review and meta-analysis, we evaluated these differences quantitively by assessing the effect of approved drugs on HbA1c levels. Both models

were able to predict the direction of the effect seen in humans for all approved drugs with more than two comparisons. Although effect sizes were larger in ZDF rats compared to db db mice for all drugs, we only found this difference to be significant for exenatide (P = 0.02). Regarding drug classes, the results are similar: only GLP-1 agonists were significantly different, primarily driven by exenatide (P = 0.03). Since such dissimilarities were not found for liraglutide, it is unlikely that they are caused by class-specific effects. It is possible ZDF rats are more sensitive to exenatide (drug-specific effect), but additional experiments are needed for confirmation.

In the db/db mouse – but not in the ZDF rat, a higher SMD was calculated for publications which started the treatment earlier and had longer treatment durations. These results could be explained by the higher use of thiazolidinediones – the drug class with more than one drug with the highest effect size – in all categories with larger effect sizes. We, therefore, conducted a sensitivity analysis and demonstrated that when the thiazolidinediones are excluded, all the above-mentioned subgroup differences disappear.

The results from this meta-analysis suggest the conclusions from FIMD's pilot study are accurate as the db/db mouse and ZDF rat were comparable for almost all drugs and classes. As a follow-up of FIMD, we designed this systematic review to lay the foundation for the development of a methodology that correlates effect sizes in pre- and clinical studies. By determining effect sizes across drugs and drug classes, we can generate tables that allow the evaluation of the similarity between these results and clinical data. A factor based on the degree of overlap between point estimates and confidence intervals could be calculated for each drug, class, and finally, animal model. Researchers could then base the selection of the model(s) for preclinical development on the extent of the human-animal overlay for new drugs that at least partially share pathways with approved drugs.

Nonetheless, for these comparisons to be scientifically valid, it is necessary to solve some methodological questions. For instance, some clinical studies compare the difference between endpoint and baseline HbA1c, and meta-analysis of generally homogeneous clinical studies calculate the Weighted Mean Difference (WMD) instead of the SMD. Further considerations about how to interpret the overlap in confidence intervals, sample size, reproducibility, species differences, dosing, pharmacokinetics and risk of bias (especially blinding and randomisation) will be crucial.

Limitations

The results of a systematic review and meta-analysis are as robust as the data sources it uses. As evidenced in our risk of bias assessment – and corroborated by

A systematic review and meta-analysis of two animal models of Type 2 Diabetes

previous literature – animal research is still poorly reported, and often essential details regarding the used methodology are not reported (29–31). Consequently, for most parameters, it was not possible to assess the risk of bias reliably. While poor reporting does not equate to poor conduct, the absence of information regarding the study design and execution prevents a rigorous evaluation of the robustness of the included studies. As such, the results of this review must be interpreted with caution.

We included only approved drugs, which have been proven to be effective in humans, naturally skewing the results towards models being considered more predictive. Ideally, we should include both effective and ineffective drugs since a predictive model should also simulate the absence (or opposite direction) of response. Nevertheless, due to the publication bias common in animal research, most of these studies are not published and therefore, would not be included in the review regardless.

As suggested by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) for animal studies, we also considered the indirectness, inconsistency, imprecision and publication bias (32). In terms of indirectness, our research included mice and rats, which are inferior species that have relevant known and unknown disparities when compared to humans. We selected the same interventions and outcomes used in clinical research, partially offsetting these concerns, but durations, routes of administration, doses, age and other factors regularly differ from clinical practice.

As for inconsistency, heterogeneity levels were generally high, which is not very surprising as animal studies are often explorative and heterogeneous regarding design and intervention protocols when compared to clinical trials. Exploring this heterogeneity is one of the added values of meta-analyses of animal studies and it might help to inform the design of future preclinical research and subsequent clinical trials. However, to account for the anticipated heterogeneity, we used a random- rather than fixed-effects meta-analysis. The sensitivity analysis also helped to reliably interpret our subgroup results, preventing that often spurious effects be mistaken by actual effects. For example, when we excluded the thiazolidinediones in our sensitivity analysis, the subgroup differences in the db/db mouse were no longer significant.

The low number of studies for many drugs and drug classes in both models may have impaired the precision of the effect sizes reported in this meta-analysis. Finally, the publication bias assessment did not find any evidence of publication bias, contrary to what is often reported in the literature (33,34). However, most drugs and many drug classes had fewer than 10 comparisons and could not be assessed for one or both models.

Conclusion

We conducted the first systematic review and meta-analysis of animal models of type 2 diabetes, aiming at discriminating between models. The meta-analysis indicates both models respond similarly in terms of Hb1Ac reduction across drugs and drug classes, except for exenatide, which seemed to have a more substantial effect in ZDF rats. For drugs with more than two comparisons, the findings are in line with the clinical literature. These results corroborate previous research in showing the differences between the db/db mouse and the ZDF rat are unlikely to be pertinent for preliminary efficacy testing.

Acknowledgements

References

1. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov. 2004;3(8):711–5.
2. Schulz JB, Cookson MR, Hausmann L. The impact of fraudulent and irreproducible data to the translational research crisis - solutions and implementation. J Neurochem. 2016;139 Suppl 2:253–70.
3. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014 Jan;32(1):40–51.
4. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. JAMA Intern Med. 2016 Dec 1;176(12):1826–33.
5. OCDE/EU. Health at a Glance: Europe 2018: State of the Health in the EU Cycle [Internet]. OECD Publishing; 2018 [cited 2019 Oct 14]. Available from: https://doi.org/10.1787/health_glance_eur-2018-en
6. International Diabetes Federation. IDF Diabetes Atlas, 8th [Internet]. International Diabetes Federation; 2017 [cited 2019 Oct 14]. Available from: http://www.diabetesatlas.org
7. Ferreira GS, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, Schellekens H, et al. Correction: A standardised framework to identify optimal animal models for efficacy assessment in drug development. PLoS ONE. 2019 Jul 22;14(7):e0220325.
8. S. Ferreira G, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, Schellekens H, et al. A standardised framework to identify optimal animal models for efficacy assessment in drug development. Macleod MR, editor. PLoS ONE. 2019 Jun 13;14(6):e0218014.
9. Wang B, Chandrasekera PC, Pippin JJ. Leptin- and leptin receptor-deficient rodent models: relevance for human type 2 diabetes. Curr Diabetes Rev. 2014 Mar;10(2):131–45.
10. Nathan DM, Turgeon H, Regan S. Relationship between glycated haemoglobin levels and mean glucose levels over time. Diabetologia. 2007 Nov;50(11):2239–44.
11. Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ, et al. Translating the A1C assay into estimated average glucose values. Diabetes Care. 2008 Aug;31(8):1473–8.
12. FDA. Guidance for Industry Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention [Internet]. Food and Drug Administration; 2008 [cited 2019 Oct 14]. Available from: https://www.fda.gov/media/71289/download
13. EMA. Guideline on clinical investigation of medicinal products in 5 the treatment or prevention of diabetes mellitus [Internet]. European Medicines Agency; 2018 [cited 2019 Oct 14]. Available from: https://www.ema.europa.eu/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-treatment-prevention-diabetes-mellitus_en.pdf
14. Hummel KP, Coleman DL, Lane PW. The influence of genetic background on expression of mutations at the diabetes locus in the mouse. I. C57BL/KsJ and C57BL/6J strains. Biochem Genet. 1972 Aug;7(1):1–13.
15. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016 05;5(1):210.
16. Rohatgi A. WebPlotDigitizer [Internet]. 2019 [cited 2019 Jan 8]. Available from: https://automeris.io/WebPlotDigitizer
17. Hooijmans CR, Rovers MM, de Vries RBM, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol. 2014 Mar 26;14:43.
18. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003 Sep 6;327(7414):557–60.
19. The Jackson Laboratories. MOUSE STRAIN DATASHEET - 000642 [Internet]. MOUSE STRAIN DATASHEET - 000642. [cited 2019 Sep 12]. Available from: https://www.jax.org/strain/000642

20. The Zucker Diabetic Fatty (ZDF) Rat: Diet Evaluation Study for the Induction of Type 2 Diabetes in Obese Female ZDF Rats [Internet]. [cited 2019 Sep 12]. Available from: https://www.criver.com/sites/default/files/Technical%20Resources/Diet%20Evaluation%20Study%20for%20the%20Induction%20of%20Type%202%20Diabetes%20in%20Obese%20Female%20ZDF%20Rats.pdf

21. Jiang B, Guo S, Shi D, Guo C, Wang T. Discovery of novel bromophenol 3,4-dibromo-5-(2-bromo-3,4-dihydroxy-6- (isobutoxymethyl)benzyl)benzene-1,2-diol as protein tyrosine phosphatase 1B inhibitor and its anti-diabetic properties in C57BL/KsJ-db/db mice. European Journal of Medicinal Chemistry. 2013;64:129–36.

22. Peterson RG. alpha-Glucosidase inhibitors in diabetes: lessons from animal studies. Eur J Clin Invest. 1994;24:11–8.

23. Tullin S, Sams A, Br, t J, Dahl K, Gong W, et al. Recombinant Adiponectin Does Not Lower Plasma Glucose in Animal Models of Type 2 Diabetes. PLoS ONE [Internet]. 2012;7(10). Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L365767527 http://dx.doi.org/10.1371/journal.pone.0044270 http://www.plosone.org/article/fetchObjectAttachment.action?uri=info%3Adoi%2F10.1371%2Fjournal.pone.0044270&representation=PD

24. Dhuria RS, Singh G, Kaur A, Kaur R, Kaur T. Current status and patent prospective of animal models in diabetic research. Adv Biomed Res. 2015;4:117.

25. Ainge H, Thompson C, Ozanne SE, Rooney KB. A systematic review on animal models of maternal high fat feeding and offspring glycaemic control. Int J Obes. 2011 Mar;35(3):325–35.

26. Saleh F, Itani L, Calugi S, Grave RD, El Ghoch M. Adipose-derived Mesenchymal Stem Cells in the Treatment of Obesity: A Systematic Review of Longitudinal Studies on Preclinical Evidence. CSCR. 2018 Jul 2;13(6):466–75.

27. Ranasinghe P, Jayawardana R, Galappaththy P, Constantine GR, de Vas Gunawardana N, Katulanda P. Efficacy and safety of 'true' cinnamon (*Cinnamomum zeylanicum*) as a pharmaceutical agent in diabetes: a systematic review and meta-analysis: Efficacy and safety of 'true' cinnamon in diabetes. Diabetic Medicine. 2012 Dec;29(12):1480–92.

28. Varga OE, Zsíros N, Olsson IAS. Estimating the predictive validity of diabetic animal models in rosiglitazone studies: Predictive validity of diabetic animal models. Obes Rev. 2015 Jun;16(6):498–507.

29. Bebarta V, Luyten D, Heard K. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad Emerg Med. 2003 Jun;10(6):684–7.

30. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012 Mar 28;483(7391):531–3.

31. Hooijmans CR, Draper D, Ergün M, Scheffer GJ. The effect of analgesics on stimulus evoked pain-like behaviour in animal models for chemotherapy induced peripheral neuropathy- a meta-analysis. Sci Rep. 2019 Dec;9(1):17549.

32. Wei D, Tang K, Wang Q, Estill J, Yao L, Wang X, et al. The use of GRADE approach in systematic reviews of animal studies: Use GRADE in animal study. Journal of Evidence-Based Medicine. 2016 May;9(2):98–104.

33. Green SB. Can animal data translate to innovations necessary for a new era of patient-centred and individualised healthcare? Bias in preclinical animal research. BMC Med Ethics. 2015 Dec;16(1):53.

34. Sena ES, van der Worp HB, Bath PMW, Howells DW, Macleod MR. Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy. Roberts I, editor. PLoS Biol. 2010 Mar 30;8(3):e1000344.

A systematic review and meta-analysis of two animal models of Type 2 Diabetes

1) Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands;
2) Copernicus Institute of Sustainable Development, Innovation Studies, Utrecht University, Utrecht, The Netherlands;
3) Medicines Evaluation Board (CBG), Utrecht, The Netherlands;
4) Centre for Human Drug Research, Leiden, the Netherlands
5) Central Committee on Research Involving Human Subjects (CCMO), Leiden, The Netherlands
6) Leiden University Medical Centre, Leiden, the Netherlands

5

Guilherme S. Ferreira[1],
Désirée H. Veening-Griffioen[1],
Wouter P.C. Boon[1,2],
Huub Schellekens[1],
Ellen H. M. Moors[1,2],
Peter J.K. van Meer[1,3],
Frederik E. Stuurman[4,6],
Joop M.A. van Gerven[4,5,6]
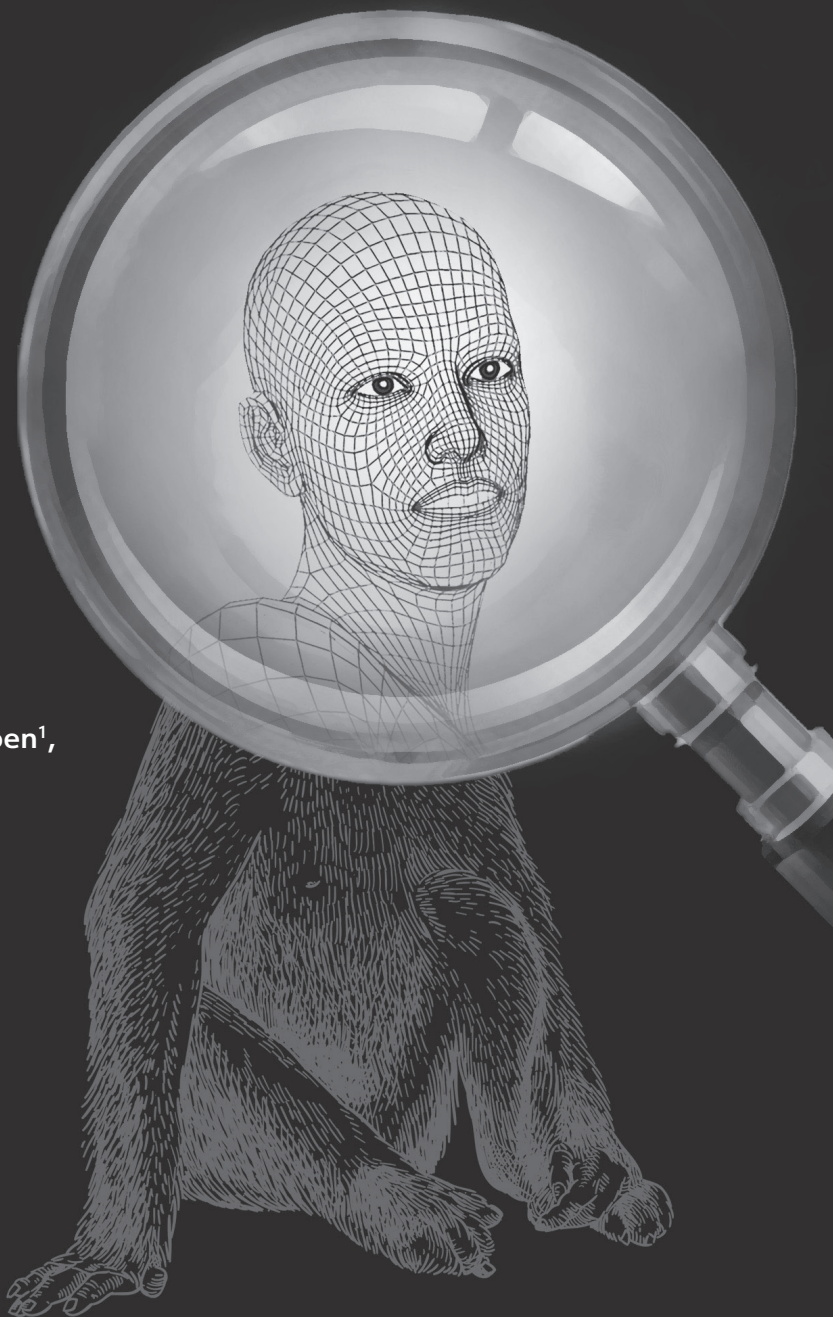
THE INVESTIGATOR'S BROCHURE (IB) AND THE TRANSLATIONAL RELEVANCE OF PHARMACOKINETIC PARAMETERS

Abstract

Insufficient efficacy and safety are the main reasons for failure in clinical development. The lack of predictivity of animal studies is often cited as a potential explanation to the attrition in the clinical stage. We use Investigator's Brochures to explore the predictivity of animal pharmacokinetic and pharmacodynamic parameters. Animal model justification, internal validity and reporting quality data of the animal experiments were collected from the IBs. We compared the preclinical and clinical Human Equivalent Dose (HED), maximum concentration in plasma (C_{max}) and Area Under the Curve (AUC) of the concentration-time curve ranges of phase I trials, and explored how these related to preclinical pharmacological effects. The IB-derisk tool was used to integrate the preclinical and clinical pharmacokinetic parameters. We analysed pharmacologically active and toxicological ranges and calculated the overlap with clinical data, determined from the early clinical studies with the compounds. Most animal studies in the IBs did not report the sex, strain or route of administration of all experiments, and none reported any measure to reduce bias (e.g. randomisation, blinding). Less than one third of the IBs included a justification for the animal model choice. In safety, the actual human HED, Cmax and AUC did not exceed the lowest No-Adverse Event Level (NOAEL) reported in the IB. Cmax predicted the congruency of safety ranges between animals and humans most accurately and regardless of indication. In efficacy, overlaps between animal and human data were slightly lower in C_{max} (65%) than in HED (70%) and AUC (77%). The overlaps were similar for various CNS drugs (HED: 81%, C_{max}: 65%, AUC: 80%) but different for non-CNS drugs (HED: 48%, C_{max}: 67%, AUC: 71%). Our results suggest that while animal models can potentially predict human well-tolerated and pharmacologically active ranges, there is much room for improvement. The main reasons for small overlaps were inadequate endpoints, different routes of administration, poor understanding of the mechanism of action and unaccounted species differences. We demonstrate that the IB-derisk tool can be used to integrate animal and human pharmacokinetic data to assess the predictivity of animal models. With a more detailed translational rationale, better reporting and more extensive pharmacokinetic characterisation of animal efficacy, the IB can be used as a powerful translational tool.

Keywords: animal model; drug development; pharmacokinetics; translational research; clinical pharmacology; investigator's brochure

Introduction

The lack of safety and efficacy are the main reasons for failure in the clinical stage of drug development (1–4). The translational challenges faced in the transition to human testing led to the definition of a 'valley of death', which encompasses the early development up to phase II clinical trials (5). The lack of predictivity of animal studies has been pointed out as a significant contributor to such a high failure rate (6). Only recently, more attention has been directed towards the external validity of animal models (7,8). However, how we can make this assessment at the various stages of development remains underexplored.

The Investigator's Brochure (IB) is a document required by the Good Clinical Practice (GCP) guidelines of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). It compiles nonclinical and clinical information on investigational products relevant for their study in humans (9). The ICH-GCP guidelines present the minimum requirements of the IB, although, in practice, the content may vary significantly (9). The IB is the primary source for the development of the translational strategy, which allows the investigators to assess the risk-benefit of the proposed trial independently. For first-in-human (FIH) studies, it also includes the translation of data from animal models to humans, including the proposed dose range. Although current guidelines suggest the consideration of pharmacological effects seen in preclinical studies, the No Observed Adverse Event Level (NOAEL) is often the basis for dose selection in first-in-human trials (10–12).

The translation between animal and human doses considers three key pharmacokinetic (PK) parameters: the Human Equivalent Dose (HED), the maximum plasma concentration (C_{max}) and the exposure, known as Area Under the Curve (AUC) of the concentration-time curve. The HED is frequently calculated based on the dose (mg/kg) and the body surface area relationship between species, but it can also use the body weight (10,13). Data on HED, C_{max} and AUC are highly heterogeneous across IBs, and their translatability across species is uncertain. Hence, the integration of PK parameters into the overall clinical development plan can be challenging (11).

To facilitate this integration, the Centre for Human Drug Research (CHDR) has developed a tool to integrate preclinical and clinical data based on the HED, C_{max} and AUC – the IB-derisk tool, available at https://www.ib-derisk.org/login/ (11). The IB-derisk tool allows the extrapolation of missing PK parameters (which are not gathered in every preclinical experiment) and comparison of the results of animal studies with predicted or collected human data. The IB-derisk tool uses colour-codes to enable the quick identification of dose-response patterns of pharmacological and toxicological effects.

Our research assesses the accuracy of the HED, C_{max} and AUC to predict pharmacological and toxicological ranges in humans integrating the data of the IBs using the IB-derisk tool. Also, we investigate the justification for animal model selection and the quality of the preclinical evidence (internal validity and reporting quality) presented in each IB.

Methods

Investigator's Brochure selection

IBs from finished clinical studies conducted from 2009 to 2019 at CHDR were retrieved from the internal database. Identities of compounds were blinded. We included IBs of clinical studies with at least two dose levels that assessed and reported statistically significant results of any measure of pharmacological activity related to a drug's mechanism of action. We excluded trials testing combinations and drugs already approved by the Food and Drug Administration (FDA) or the European Medicines Agency (EMA). For these products, the Summary of Product Characteristics (SmPC) is usually provided to the investigator instead of an IB and, therefore, animal pharmacokinetic and pharmacodynamic data are not systematically reported. An IB-derisk assessment was conducted for all included IBs according to our previous publication (11).

Data collection

From the IBs, we extracted the therapeutic area of the proposed indication; the mechanism of action; the HED, C_{max} and AUC from the lowest and highest doses for which a pharmacological effect related to efficacy was reported; anticipated therapeutic dose and its rationale; the lowest and highest NOAELs across species and the HED, C_{max} and AUC associated with them; internal validity information (randomisation, blinding); reporting quality (strain, sex, route of administration); justification for the use of the models of disease. The HED was calculated according to the factors provided by the FDA guidelines (10). The clinical data were extracted from protocols and clinical study reports (CSRs) of the corresponding studies at CHDR and complemented with the clinical data provided in the IB whenever available.

Parameter inter- and extrapolation

Since most pharmacokinetic parameters are often not available for animal studies, we inter- or extrapolated accordingly, as described previously (11). Whenever at least one C_{max} or AUC value was available for a dose level in a certain species, this value was replicated in other experiments in the same species at the same dose level, if C_{max} or

AUC was not reported. If more than one value was available for the same dose level, we calculated the average of the parameter. If no values were available for a dose level, we interpolated the C_{max} and AUC using the dose levels immediately below and above. If only dose levels above or below the dose level with a missing value were available, we calculated it proportionally to the available values. Values were inter- or extrapolated taking species into account, and sex or strain whenever available.

Data analysis

For each parameter (HED, C_{max} and AUC), we calculated the percentage overlap of the preclinical dose range based on the clinical range for efficacy pharmacological effects. Here, we use 'efficacy' to describe pharmacological effects that are expected to be associated with therapeutic effects. The preclinical and clinical efficacy dose range was defined by the lowest and the highest dose in which any pharmacological effect associated with the proposed mechanism of action of the drug was reported. The preclinical safety dose range was determined by the lowest and the highest NOAEL across species. In humans, it was determined by the range in which the drug was well-tolerated, i.e. no severe adverse reactions or permanent discontinuation due to a drug-related adverse event was reported. For safety, we evaluated whether the clinical range was inferior to the pharmacokinetic parameters associated with the lowest and highest NOAELs reported. If it was superior, we calculated the degree to which the clinical range surpassed the animal levels.

Results

IB General Characteristics

The search in the CHDR's database from 2009 to 2019 returned 187 finished clinical trials and their corresponding IBs. Of these, 25 trials met the inclusion criteria. Table 1 shows the reasons for exclusion of the remaining 162. All IBs were from phase I pharmacodynamics studies, of which 14 (62%) were FIH. Data from clinical studies provided in the IBs were also included in our analysis. Most IBs described drugs developed for central nervous system (CNS) (68%, N = 17/25) and endocrine/metabolic (16%, N = 4/25) conditions. The other four IBs were for respiratory/sensory organ, urogenital, autoimmune and musculoskeletal disease, all with one compound each. The anticipated therapeutic dose was reported in the IB or protocol for most drugs (84%, N = 21/25). Most IBs included considerations on both safety and efficacy in the dosing rationale (72%, N = 18/25). All studies with a dosing rationale that considered only the NOAEL were first-in-human. Table 2 describes the IB characteristics.

TABLE 1 | Reasons for exclusion of IBs and number of IBs excluded.

Reason for exclusion	N
Non-interventional	39
Registered drug	31
No PD data	27
One dose	13
Not available	12
No PK data	12
Drug-Drug Interaction	9
Safety	8
No effective dose	5
Not conducted	3
Combination	1
Drug not in development	1
Medical Device	1
Total	162

CSR – clinical study report; PK – pharmacokinetics; PD - pharmacodynamics

Animal model justification, internal validity and reporting quality

Most IBs included animal models of disease or testing paradigms (e.g. Morris Water Maze) in their preclinical studies (84%, N = 21/25), but only a few justified the model selection in some way (28%, N = 7/25). The motivations included availability (commonly used models), similar phenotype (symptoms), response to effective drugs (pharmacology), histology and biomarkers. There was no explanation as to the relevance of model choice when compared to other available options. Although all animal experiments included a placebo arm, blinding or randomisation was not reported in any. The strain and sex were frequently missing. Only one IB reported the strain, sex and route of administration for all studies. In the studies that did report the sex, male animals were overrepresented (data not shown).

Safety Assessment

Table 3 shows the safety pharmacokinetic ranges and the lowest and highest reported NOAELs per IB. As shown in Figure 1, the highest human well-tolerated dose's HED, C_{max} and AUC surpassed the lowest NOAEL's in 40% (N = 10/25), 16% (N = 4/25) and 36% (N = 9/25) of drugs, respectively. Only in one trial was the highest NOAEL's HED exceeded (IB19). This exploration of supratherapeutic doses was intended to investigate the drug's abuse potential per FDA's guideline "Assessments of Abuse Potential of Drugs" (14). The medians were 100% across all three parameters for the lowest and highest NOAELs. When analysed per indication (CNS vs non-CNS), the trend was similar for CNS drugs only – the medians for HED (N = 17), C_{max} (N = 17) and AUC (N = 16) were all 100%. In non-CNS drugs, the medians were 109% (HED), 100% (C_{max}), and 188% (AUC).

TABLE 2 | Therapeutic area, phase, first-in-human status, anticipated dose, presence of animal models of disease or paradigms, model motivation, animal studies internal validity and reporting quality of the included IBs.

IB	Therapeutic Area	Mechanism of Action	FIH? (Y/N)	Anticipated Therapeutic Dose (mg)	Dosing Rationale	DMP	Model Motivation
1	Central Nervous System	GABA_A-α2,3R partial agonist	Y	29	Safety	Y	Commonly used Pharmacological
2	Central Nervous System	GFRα3 co-receptor ligand	N	0.8	Efficacy + Safety	Y	Phenotype
3	Central Nervous System	OX1R and OX2R antagonist	Y	NR	Safety	N	NR
4	Central Nervous System	H3R agonist	Y	NR	Safety	Y	NR
5	Central Nervous System	H3R antagonist	N	5	Efficacy + Safety	Y	NR
6	Central Nervous System	GABAα receptor ligand	Y	0.4	Efficacy + Safety	Y	NR
7	Central Nervous System	α7 nAChR agonist	N	10	Efficacy + Safety	Y	NR
8	Central Nervous System	OX2R antagonist	N	10	Efficacy + Safety	Y	NR
9	Central Nervous System	TAAR1 partial agonist	N	180	Efficacy + Safety	Y	NR
10	Central Nervous System	OX1R and OX2R antagonist	Y	NR	Safety	N	NR
11	Central Nervous System	P2X7R antagonist	N	10	NR	Y	NR
12	Central Nervous System	M1R partial agonist	Y	120	Safety	Y	Histology
13	Central Nervous System	nAChR inhibitor (prodrug)	Y	NR	Safety	Y	Biomarker
14	Central Nervous System	OX1R antagonist	Y	4	Safety + Efficacy	Y	NR
15	Central Nervous System	GCase allosteric activator	Y	24.4	Safety	Y	NR
16	Central Nervous System	GABA_A α5 positive allosteric modulator	Y	65 BID	NR	Y	NR
17	Central Nervous System	AMPAR positive allosteric modulator	N	0.5	Efficacy + Safety	Y	NR
18	Respiratory/Sensory Organ	Fibrosis inhibitor	N	1	Efficacy + Safety	Y	Histology
19	Urogenital	FAAH inhibitor	N	600	Efficacy + Safety	Y	NR
20	Endocrine-Metabolic	sst2R and D2R agonist	Y	0.11	Efficacy + Safety	Y	Commonly used
21	Endocrine-Metabolic	Amylin- and Calcitonin-Receptor Agonist	Y	0.01	Efficacy + Safety	Y	NR
22	Endocrine-Metabolic	Amylin- and Calcitonin-Receptor Agonist	Y	0.005	Efficacy + Safety	Y	Symptomatology
23	Immunology	anti-OX40 ligand (OX40L) subclass G4PE kappa	Y	0.42	Efficacy + Safety	Y	NR
24	Musculoskeletal	p38α/β inhibitor	N	7.5 BID	Efficacy + Safety	N	NR
25	Endocrine-Metabolic	GLP-2R agonist	N	1	Efficacy + Safety	N	NR

DMP – animal model of disease or paradigm; FIH – first-in-human; N – No; NR – not reported; PR – partially reported; Y – yes.

| The Investigator's Brochure (IB) and the translational relevance of pharmacokinetic parameters

TABLE 3 | Animal and human safety dose ranges per pharmacokinetic parameter.

IB	L-NOAEL (mg/kg)	H-NOAEL (mg/kg)	Human Equivalent Dose (mg)						C_{max} (ng/ml)		
			ALR	AUR	HLR	HUR	L-Overlap (%)	H-Overlap (%)	ALR	AUR	HLR
1	10	250	97.2	8,115	10	600	617	100	2,310	60,500	9.4
2	4	8	77.8	155.5	0.4	0.8	100	100	134	3,511	33.2
3	100	1,000	972	32,460	5	1,500	154	100	1,817	16,450	40.1
4	1	15	32.46	148.5	0.25	1.5	100	100	609	3,440	0.22
5	3	300	58.3	2916	0.005	0.125	100	100	610	42,435	0.00913
6	6	10	58.3	324.6	0.04	100	172	100	1,230	6,350	0.229
7	50	2,000	1,623	19,440	10	80	100	100	700.5	2,460.7	7.8
8	8.25	1,000	267.8	9,720	10	40	100	100	1,270	35,143	183
9	45	200	874.8	1,944	30	450	100	100	4,790	4,820	66.3
10	20	1,000	649.2	9,720	5	200	100	100	2,020	11,928	159.7
11	15	250	486.9	2,430	50	450	100	100	1,350	10,005	138
12	1	300	32.5	2,916	1	35	108	100	297	37,700	3.4
13	2	10	19.4	97.2	11	44	227	100	27.1	134	10.7
14	50	150	486	4,869	1	90	100	100	6,976	27,400	97.4
15	25	200	811.5	1,944	3	30	100	100	6,315	59,700	110
16	80	450	777.6	2,916	2.5	375	100	100	5,960	43,300	32.5
17	0.3	200	5.83	1,944	0.3	18	309	100	148	2,678	2.93
18	4	200	16.2	2,332.8	60	600	3,704	100	374,000	4,447,000	27,200
19	10	100	97.2	1,166.4	1	1,800	1,852	154	7,280	96,300	17
20	0.1	1	0.97	19.44	0.1	2	206	100	4.5	625	0.707
21	0.05	0.05	0.486	0.05	0.005	0.04	100	100	61.416	NA	0.031
22	0.0175	0.025	0.568	0.812	0.005	0.025	100	100	10.45	13.59	0.08
23	50	100	972	1,944	0.4	360	100	100	2,445,000	3,025,000	130
24	3	1,000	29.2	19,440	7.5	7.5	100	100	712	1,087	36.6
25	5	250	48.6	2430	1	56.9	117	100	4,500	50,000	13.9
Median							100	100	Median		

ALR – Animal Lower Range; AUR – Animal Upper Range; HLR – Human Lower Range; HUR – Human Upper Range; H-overlap – overlap with the highest NOAEL; L-overlap – overlap with the lowest NOAEL; NA – not applicable; NR – not reported; NOAEL – No-Observed Adverse Event Level

| Cmax (ng/ml)			AUC (ng.h/ml)						
HUR	L-Overlap (%)	H-Overlap (%)	ALR	AUR	HLR	HUR	L-Overlap (%)	H-Overlap (%)
1,430	100	100	29,800	201,293	9	5,790	100	100
156	116	100	1,763	2,294	40.7	606	100	100
868.9	100	100	9,080	130,500	161.4	26,585.8	293	100
1.4	100	100	756	15,826	0.29	3.55	100	100
0.245	100	100	7,584	435,485	0.245	3.925	100	100
559.3	100	100	9,850	10,100	1.6	5,753	100	100
60.2	100	100	5,906	14,329.3	131.7	1,031.4	100	100
570	100	100	3,570	135,927	982	3,062	100	100
1,130	100	100	33,700	63,150	743	14,600	100	100
1,868.9	100	100	10,500	142,950	971.3	22,906.4	218	100
1,180	100	100	22,750	199,500	2,456	22,357	100	100
235	100	100	1,960	284,500	42	2,210	113	100
58.5	216	100	34.8	119	15.7	119	342	100
4,575	100	100	NR	NR	1,969	64,642	NA	NA
1,510	100	100	73,908	691,000	1,600	48,400	100	100
3,520	100	100	91,700	472,000	286	46,500	100	100
126	100	100	2,093	21,421	47.7	8,882	424	100
231,000	100	100	1,606,000	15,867,000	405,000	4,440,000	276	100
62,930	864	100	8,104	383,000	14	235,725	2909	100
19.4	431	100	8	1,639	0.582	41.2	515	100
0.238	100	100	33.055	NA	0.0844	0.499	100	100
0.24	100	100	18.7	24.022	0.07	0.37	100	100
270,000	100	100	26,650,000	179,000,000	26,210	169,810,000	637	100
85	100	100	4,887	32,800	26.9	42.3	100	100
1,167.13	100	100	163,000	4,465,000	35.2	96,889	100	100
Median	100	100	**Median**				100	100
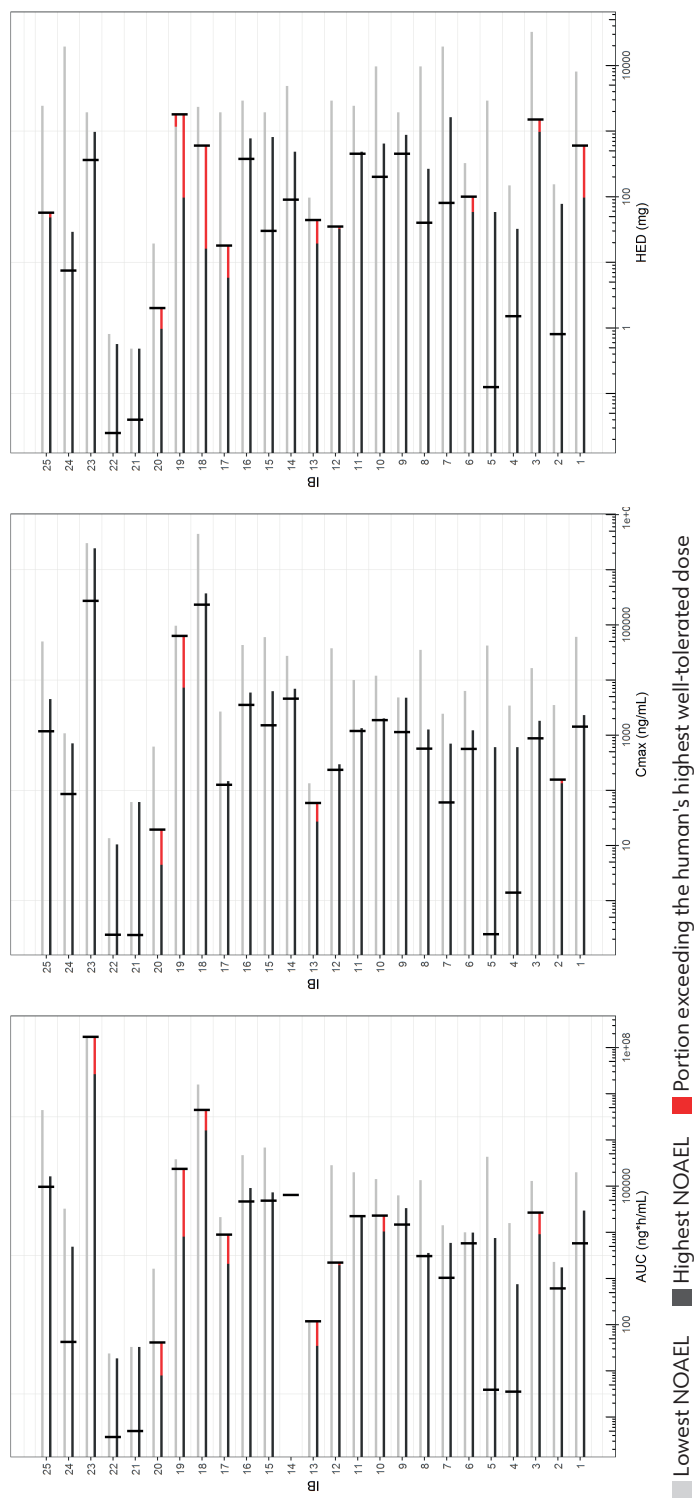
FIGURE 1 | Human Equivalent Dose (HED), maximum concentration in plasma (C$_{max}$) and Area Under the Curve (AUC) of the concentration-time curve associated to the lowest and highest No-Adverse Event Levels (NOAELs). For each IB, the vertical line represents the highest well-tolerated dose tested in humans. The portion of the lowest and highest NOAELs that exceed the highest well-tolerated dose in humans is coloured red.

Efficacy Assessment

We use 'efficacy assessment' as the assessment of pharmacological effects expected to be associated with therapeutic effects. Table 4 shows the dosing type, and pharmacokinetic efficacy ranges for HED, C_{max} and AUC per IB. The HED (70%, N = 25) and AUC (73%, N = 23) were slightly better for predicting pharmacologically active ranges than C_{max} (65%, N = 25). The medians indicate a similar interpretation (95%, 93% and 100%, respectively). Figure 2 shows the HED, C_{max} and AUC overlaps per IB. Overall, preclinical data either predicted the active ranges to a high (≥ 80% overlap in 60%, 64% and 68% of IBs based on HED, C_{max} and AUC, respectively) or low degree (≤ 20% overlap in 20%, 28% and 16% of IBs based on HED, C_{max} and AUC, respectively).

When splitting per therapeutic area, the three parameters tell a different story. In CNS, while the C_{max} overlap remained unchanged (65%, N = 17), both HED (N = 17) and AUC (N = 15) average overlaps improved to 80% and 79%, respectively, increasing the difference to C_{max} by almost twofold. The HED and AUC medians increased to 100%, while C_{max} decreased to 88%. For non-CNS indications (N = 8), the C_{max} overlap increased slightly to 67% (median: 98%), but HED and AUC overlaps decreased to 48% (median: 44%) and 61% (median: 94%), respectively.

Most IBs determined the lowest effective dose in animals, meaning the lowest dose tested was ineffective (64%, N = 16/25). In the remaining 36%, the lowest tested dose was also found to be effective, and therefore, the lowest effective dose was not necessarily established preclinically. In two of these, the lowest dose elicited an effect in some but not all experiments. Conversely, in most clinical trials, the lowest tested dose was also effective (60%, N = 15/25). In one case, the same dose was both effective and ineffective in humans across different trials (IB15).

The Investigator's Brochure (IB) and the translational relevance of pharmacokinetic parameters

TABLE 4: Animal and human efficacy dose ranges per pharmacokinetic parameter.

IB	Lowest Dose Type		Human Equivalent Dose (mg)					
	Animal	Human	ATD (mg)	ALR	AUR	HLR	HUR	Overlap (%)
1	Effective dose	Effective dose	29	2.9	29.16	300	600	100
2	Effective dose	Effective dose	0.8	2.4	16,230	0.4	0.8	100
3	Effective dose	Effective dose	NR	0.243	972	200	1,500	93
4	Effective dose	Effective dose	NR	97.2	9,720	1.5	5	100
5	Effective dose	Tested effective dose	5	97.2	291.6	0.005	5	100
6	Effective dose	Tested effective dose	0.4	64.9	2,916	0.04	100	100
7	Effective dose	Tested effective dose	10	0.024	3246	1	180	15
8	Effective dose	Tested effective dose	10	2.9	29.16	10	80	100
9	Effective dose	Effective dose	180	2.4	16,230	100	600	100
10	Tested effective dose	Effective dose	NR	0.243	972	25	200	59
11	Effective dose	Tested effective dose	10	97.2	9,720	50	450	49
12	Effective dose	Tested effective dose	120	97.2	291.6	3	35	0
13	Effective dose	Tested effective dose	NR	64.9	2,916	5.5	44	100
14	Tested effective dose	Effective dose	4	0.024	519.4	25	90	100
15	Tested effective dose	Tested effective dose*	24.4	4.9	291.6	3	60	63
16	Tested effective dose*	Effective dose	65 BID	24.3	218.7	20	375	92
17	Tested effective dose	Effective dose	0.5	48.6	39,600	0.5	6	100
18	Effective dose	Tested effective dose	1	0.29	486	60	600	0
19	Effective dose	Tested effective dose	600	0.162	38.9	1	1,800	5
20	Tested effective dose	Tested effective dose	0.11	9.7	97.2	0.1	2	100
21	Effective dose	Tested effective dose	0.01	0.019	1,166	0.005	0.04	46
22	Tested effective dose*	Tested effective dose	0.005	0.024	19.44	0.005	0.025	95
23	Effective dose	Effective dose	0.42	0.006	2.921	13.5	360	42
24	Tested effective dose	Tested effective dose	7.5 BID	48.6	194.4	7.5	15	0
25	Tested effective dose	Effective dose	1	0.012	2.9	5	56.9	100
Average								70
Median								95

* - Dose not consistently associated with pharmacological effects; ALR – Animal Lower Range; ATD – Anticipated Therapeutic Dose; AUR – Animal Upper Range; HLR – Human Lower Range; HUR – Human Upper Range;

|	Cmax (ng/ml)					AUC (ng.h/ml)				
	ALR	AUR	HLR	HUR	Overlap (%)	ALR	AUR	HLR	HUR	Overlap (%)
	51.9	15,619	477	1,430	100	396.5	250,938	1,440	5,790	100
	0.177	1,5855	149.3	156	100	2.82	14,523	347	606	100
	1660	16,450	479.3	868.9	0	6,741	130,500	6,113.1	26,585.8	72
	2.1	700	1.4	7.1	88	0.6	1,189	3.55	14.97	100
	0.074	42,435	0.009	1.144	94	0.205	435,485	0.245	262.66	100
	87.2	17,000	0.229	559.3	84	775	122,000	1.6	5,753	87
	0.342	3.42	0.572	98.8	3	1.014	10.14	12.1	8,666	0
	169.8	26,371	183	1,208	100	297	2,000,000	982	3,648	100
	429.7	4,200	261	1,660	88	2,180	39,007	2,750	19,800	100
	239	11,113	631.58	1,868.9	100	290	166,600	2,703.2	22,906.4	100
	1,214	4,013	138	1,180	0	17,909	74,330	2,456	22,357	22
	523	41,300	13.7	235	0	3,620	301,000	134	2,291	0
	3.3	1,740	10.7	58.5	100	1.7	2,990	15.7	119	100
	9.6	2,698	2,230	4,575	0	NR	NR	32,831	64,642	NA
	1,484	19,286	110	2,820	49	NR	NR	1,600	18,700	NA
	496	14,000	271	3,520	93	3,780	212,000	2,790	46,500	98
	1.9	883	4.19	45.99	100	8.28	10,800	181	839	100
	4,285	378,500	27,200	231,000	100	28,000	7,527,000	405,000	4,440,000	100
	277	19,439	17	62,930	30	252	8,104	14	235,725	3
	0.128	51,615	0.707	19.4	100	0.209	217,905	0.582	41.2	100
	1.022	2,922	0.031	0.238	0	0.768	1,519	0.084	0.499	0
	0.087	40.8	0.08	0.24	96	0.048	70.15	0.07	0.37	100
	265,000	426,000	11,360	290,000	9	26,100,000	179,000,000	5,680,000	169,810,000	88
	22.9	232.5	36.6	85	100	283.8	851.3	201	521	0
	22	121,000	94.09	1167.13	100	495	16,400,000	9,170.1	96,889	100
	Average				65	**Average**				77
	Median				93	**Median**				100
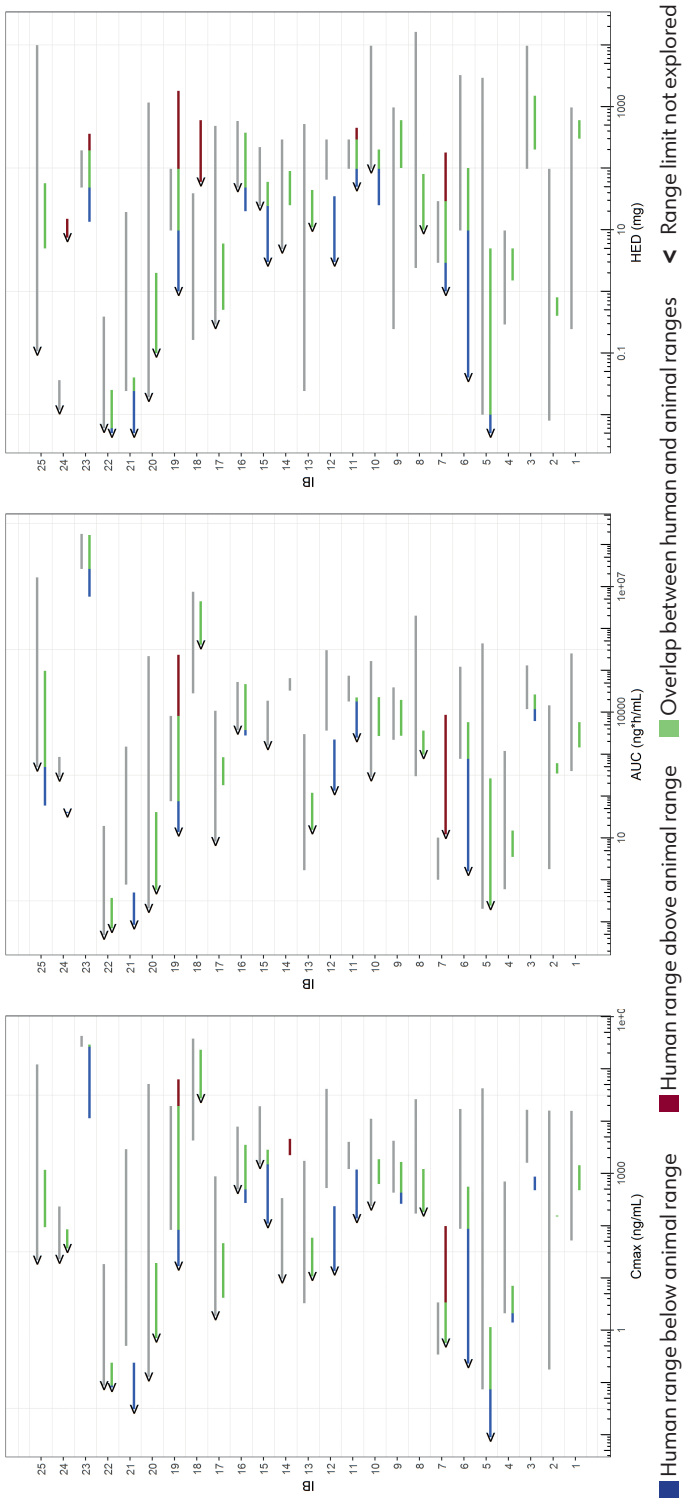
FIGURE 2 | Pharmacologically active Human Equivalent Dose (HED), maximum concentration in plasma (C_{max}) and Area Under the Curve (AUC) of the concentration-time curve animal and human ranges. The * represents the anticipated therapeutic dose. The human range is coloured blue when below and red when above animal range. The overlap is coloured green.

■ Human range below animal range   ■ Human range above animal range   ■ Overlap between human and animal ranges   ∨ Range limit not explored

Discussion

Animal model justification, internal validity and reporting quality

The IB is the first source of information for translational considerations when moving into the clinic. However, since IBs contain proprietary information without a requirement to be published, they are seldom used as data sources in scientific publications. To our knowledge, this is the first study to systematically investigate the reliability of HED, C_{max} and AUC in predicting pharmacologically active ranges in humans. To this end, the IB-derisk tool was used, which was developed to facilitate the interpretation of toxicological and pharmacological data presented in the IB. It allows the quick identification of pharmacologically active ranges and safety margins. The integration of efficacy and safety data of different species permits the assessment of therapeutic windows and the making of informed decisions on the best strategy for clinical development. We also included an assessment of how the preclinical evidence is presented, such as the motivation for model choice, internal validity and reporting quality of animal experiments.

The ICH-GCP guidelines do not formally require animal model motivation or internal validity and reporting information, leaving this at the sponsor's discretion (9). Our data corroborate previous research in showing that important details of animal studies are poorly reported not only in the literature but also in IBs (15). No IB reported blinding nor randomisation of any studies, and most lacked important information, such as animal sex or route of administration. Although IBs are meant to be a concise compilation of relevant data, this insufficient reporting is concerning given their importance in the different stages of drug development. Since poor reporting of study design is often associated with an overestimation of efficacy outcomes, it means that ethics committees and other regulatory bodies could be allowing first-in-humans trials to start on the basis of spurious results (9,15,16).

On the same note, less than one-third of IBs presented any justification for their model selection. Of those which did justify, none provided a discussion on the translational relevance of the model in comparison with other available options, including the IBs that did not include an animal model of disease or testing paradigm. The external validity of animal models has often been overlooked and, until recently, there was no standardised methodology to assess it (7,8). Such a practice may promote the use of suboptimal models without regard to the mechanism of action being investigated – generating data unlikely to translate. More importantly, it hinders a proper benefit-risk analysis of preclinical data by relevant institutions (e.g. ethics committees), promoting the advancement of clinical trials without a sound scientific rationale and therefore, unethical trials (15,17,18).

The Investigator's Brochure (IB) and the translational relevance of pharmacokinetic parameters

Safety assessment

We analysed the accuracy of HED, C_{max} and AUC with their corresponding preclinical effects, in predicting toxicological and pharmacologically active ranges in humans. For safety, our results indicate the use of the lowest NOAEL, i.e. in the most sensitive species, might lead to overly conservative estimates. Although in most human studies, the dose range was restricted by the NOAEL, several studies surpassed this level. On the other hand, the highest NOAEL predicted values much higher than the human well-tolerated tested ranges. Since most human ranges were within twofold of animal ranges and all medians were 100%, these results indicate the data were skewed by a portion of trials, mostly in CNS (e.g. IB1, IB13, IB17, IB18). First-in-human studies in CNS conduct safety dose curve explorations that may include escalations above the NOAEL if the expected adverse events can be monitored reliably (e.g. CNS-effects using the NeuroCart) (19). Additionally, since humans have slower metabolism in general, it is expected that AUC ranges may often surpass animal ranges to some extent. The NOAEL can inform human tolerability predictions most reliably if the drug's toxicity is associated to exaggerated pharmacology rather than off-target effects, which is often the case for newer, more selective compounds. The two significant outliers were actually intentional: the aforementioned IB19 was designed to explore the abuse potential of supratherapeutic doses (HED: 3704% and AUC: 276%); and IB18 comprised a recombinant endogenous protein whose pharmacological action was based on the endogenous protein's level, which was known beforehand to differ significantly between the species used in preclinical efficacy studies and humans.

The same trend was observed in a separate analysis of CNS and non-CNS drugs, in which C_{max} was the most sensitive parameter. Since C_{max} is closely associated with the volume of distribution – which does not vary highly across mammal species, it is expected to predict the incidence of acute concentration-related adverse events accurately (20). For non-CNS drugs, the HED and AUC associated with the NOAEL in the most sensitive species were relatively conservative predictors of safety, surpassing human well-tolerated ranges by 685% and 492%. Nevertheless, the extreme results for HED were mostly driven by a few IBs (median: 100%) while the median for AUC was 188%, which is acceptable as it is within a twofold range and pertains to animal-human differences in AUC. These results are indicative as the small sample sizes for non-CNS drugs (N = 8) preclude a confirmatory conclusion.

Efficacy assessment

When looking at efficacy, there is a shift in the accuracy of the exposure parameters. Both HED (70%) and AUC (77%) predicted human pharmacologically active ranges slightly more accurately than C_{max} (65%). While the C_{max} overlap was

virtually unchanged for CNS (65%) and non-CNS drugs (67%), there were larger differences in HED (80% vs 48%, respectively) and AUC (79% vs 61%, respectively). Absorption, distribution (e.g. blood-brain barrier), metabolism and excretion differences between species may partially explain the inaccurate estimation of these parameters. Further investigation is warranted on the value of HED, C_{max} and AUC across indications.

Differently from the toxicological ranges, the assessment of pharmacologically active ranges was frequently a hit or miss situation. For all three parameters, the overlaps were either large (≥ 80%) or small (≤ 20%) in at least 80% of the drugs. Small overlaps indicate inaccuracy in the predictions using preclinical data, which may result in the advancement of drug candidates with poor therapeutic potential. Although drug specificities may partially explain some small overlaps, a consistently small overlap across two or more parameters might indicate other reasons. Efficacy studies performed in some species (e.g. mouse, cat) did not explore drug exposure and therefore, no C_{max} nor AUC parameters could be inter- or extrapolated. Frequently, many toxicological studies did not include pharmacological endpoints, resulting in a significantly smaller animal range, as efficacy studies are performed at much lower dose levels. For example, in IB7, animal C_{max} and AUC corresponding to efficacy were both lower than the human counterparts owing mostly to the lack of both efficacy PK data in the mouse and pharmacological endpoints in studies exploring higher doses (e.g. toxicity). In IB21, where all three animal ranges were higher than the human ranges, the drug was administered subcutaneously while the clinical route was oral, resulting in a lower bioavailability.

Improving animal to human translation in early clinical development

These findings bring attention to the importance of the choice of animal models of disease, endpoints used in preclinical studies and determination of a dose-response curve (21,22). Although animal models are not expected to predict the human response with perfect accuracy, our results show opportunities for improvement. For example, combining safety and efficacy studies to explore a wide exposure range allows the determination of which endpoints translate across species, increasing the accuracy of pharmacokinetic predictions for humans. Disease model-specific data can also be used with the IB-derisk tool to evaluate the predictivity of particular animal models of disease. These data can be used in reverse translation research to improve suboptimal disease models.

Moreover, trials in early clinical development are designed to determine the safety, tolerability, pharmacological activity and pharmacokinetics. They are hardly ever designed to demonstrate an effect in clinical efficacy outcomes, and therefore, biochemical and physiological markers are often employed (21). The translatability

The Investigator's Brochure (IB) and the translational relevance of pharmacokinetic parameters

of the endpoints used in animal studies is rarely discussed in the IBs. As such, there are hardly any considerations on the likelihood of pharmacological effects in animals translating into actual therapeutic effects in humans. Additionally, in some cases, efficacious doses in animals may be capped by human tolerability (23).

There are other factors which may hamper a holistic assessment of the translational value of the preclinical package in the IB. Even though most animal efficacy studies are conducted in mice, pharmacokinetics is usually only marginally explored in this species. Overall, C_{max} and exposure data are either not collected or presented in the pharmacology sections, which poses a bigger challenge for the translation across species and determination of the therapeutic window. These limitations are largely related to the difficulties of PK-sampling in small laboratory animals. For instance, it was not uncommon for some routes of administration (e.g. intraperitoneal) to be used in efficacy studies without bioavailability studies across species to back them up. The employment of modelling and simulation, such as physiologically-based pharmacokinetic modelling (PBPK) before first-in-human trials, offers an opportunity for more precise estimates (24,25).

By and large, the value of IBs for the development of translational strategies can be greatly enhanced by adhering to higher reporting standards in the preclinical studies. These improvements will support the use of the IB to allow "a clinician, or potential investigator, to understand it and make his/her own unbiased risk-benefit assessment of the appropriateness of the proposed trial" (9). Such an objective assessment is only possible when the data are presented uniformly, clearly and in a proper translational context. Also, an IB in which data is transparently and accurately reported facilitates the interaction between companies and investigators, ethics committees and regulatory agencies – expediting the drug development process. The IB-derisk can be used as a tool as it orders all preclinical studies according to simple pharmacokinetic principles, and the emerging human data can be readily inserted to guide early studies in healthy volunteers and patients. The inclusion of an extensive exposure analysis of efficacy models (including the combination with toxicity studies whenever possible), analysis of the totality of the evidence in the IB, the nature of efficacy studies (exploratory or confirmatory) and detailed elaboration of the relevance of animal models and study designs offer further clear directions toward that goal (15).

Limitations

While the trends identified in our study are worth investigating, the limited number, diversity and non-randomness of the included IBs makes our findings suggestive rather than confirmatory. We designed this study to solely include IBs of drugs for which at least two (pharmacodynamically) active doses were identified in phase

I trials to allow a comparison between animal and human ranges. As such, these data cannot be used to compare the HED, Cmax and AUC regarding their ability to predict an effect in humans. Also, we could only compare the ranges reported in the IBs and CSRs from phase I studies. If dose-response curves were not fully explored, some estimations may need review, should additional data become available. In some IBs, toxicity studies reported pharmacological effects and were included in the pharmacologically active range. This inclusion resulted in much larger animal ranges, which could more easily overlap with clinical ranges in comparison to drugs without pharmacological endpoints. We used linear inter- and extrapolation to determine missing pharmacokinetic parameters in animals. Although such a strategy is common practice, it might lead to prediction inaccuracies for drugs with a non-linear pharmacokinetic profile, and PK/PD-based analyses might have been more reliable. Also, the analysis was limited to studies of unregistered compounds mostly in healthy volunteers, therefore no predictions could be made about therapeutic efficacy or clinical effects in patients.

Conclusion

This study provides a first insight into the relevance of commonly used parameters of drug exposure in laboratory animals, in the prediction of human effective dose ranges using IBs. The reporting of measures to reduce the risk of bias, basic information (e.g. strain and sex) and justification of model choice is still insufficient. Generally, IBs highlight common practices in the industry that preclude a detailed assessment by ethics committees and regulatory agencies alike, possibly resulting in clinical trials advancing on the basis of spurious results. An improvement of the reporting of study design, animal characteristics and translational rationale is long overdue.

For safety, we show that the lowest NOAEL may result in overly conservative predictions for drugs with significant off-target toxicity. The C_{max} was the most sensitive parameter to translate ranges of animal safety to human tolerability, regardless of indication. Some NOAEL-based animal HED and AUC ranges were exceeded in humans, mostly owing to reliable tools to monitor adverse events, species differences or drug specificities. For efficacy, the average overlap of the preclinical data with human pharmacologically active ranges varied from 65% (Cmax) to 77% (AUC). HED and AUC overlapped marginally more with clinical pharmacologically active ranges, but the high level of heterogeneity creates a challenge to their discrimination. The differences between CNS and non-CNS drugs warrant further investigation as they might present an opportunity for a better application of these parameters in early clinical development.

The IB-derisk tool allows a first look into commonly used pharmacokinetic parameters across a pool of drugs and indications semi-quantitatively. It can also be used to assess the predictivity of specific animal models, adding to a crescent range of new strategies to improve the preclinical translation. With better reporting and more detailed translational rationales, the IB can be used as a powerful instrument for the translation of animal to human data.

Acknowledgements

References

1. Arrowsmith J. Trial watch: phase III and submission failures: 2007-2010. Nat Rev Drug Discov. 2011 Feb;10(2):87.
2. Arrowsmith J, Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. Nat Rev Drug Discov. 2013 Aug;12(8):569.
3. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014 Jan;32(1):40–51.
4. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. JAMA Intern Med. 2016 Dec 1;176(12):1826–33.
5. Seyhan AA. Lost in translation: the valley of death across preclinical and clinical divide – identification of problems and overcoming obstacles. transl med commun. 2019 Dec;4(1):18.
6. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can Animal Models of Disease Reliably Inform Human Studies? PLoS Med. 2010 Mar 30;7(3):e1000245.
7. S. Ferreira G, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, Schellekens H, et al. A standardised framework to identify optimal animal models for efficacy assessment in drug development. Macleod MR, editor. PLoS ONE. 2019 Jun 13;14(6):e0218014.
8. Ferreira GS, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, Schellekens H, et al. Correction: A standardised framework to identify optimal animal models for efficacy assessment in drug development. PLoS ONE. 2019 Jul 22;14(7):e0220325.
9. European Medicines Agency. ICH E6 (R2) Good clinical practice - Step 5 [Internet]. 2016 [cited 2020 Apr 28]. Available from: https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice
10. Food and Drug Administration. Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers: Guidance for Industry [Internet]. 2005 [cited 2020 Apr 29]. Available from: https://www.fda.gov/media/72309/download
11. van Gerven J, Cohen A. Integrating data from the Investigational Medicinal Product Dossier/ investigator's brochure. A new tool for translational integration of preclinical effects. Br J Clin Pharmacol. 2018 Jul;84(7):1457–66.
12. European Medicines Agency. Guideline on strategies to identify and mitigate risks for first-in-human and early clinical trials with investigational medicinal products [Internet]. 2017 [cited 2020 Apr 5]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-strategies-identify-mitigate-risks-first-human-early-clinical-trials-investigational_en.pdf
13. Nair A, Jacob S. A simple practice guide for dose conversion between animals and human. J Basic Clin Pharma. 2016;7(2):27.
14. Food and Drug Administration. Assessment of Abuse Potential of Drugs: Guidance for Industry [Internet]. 2017 [cited 2020 Jun 18]. Available from: https://www.fda.gov/media/116739/download
15. Wieschowski S, Chin WWL, Federico C, Sievers S, Kimmelman J, Strech D. Preclinical efficacy studies in investigator brochures: Do they enable risk–benefit assessment? Macleod M, editor. PLoS Biol. 2018 Apr 5;16(4):e2004879.
16. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ. 2007 Jan 27;334(7586):197.
17. Kimmelman J, Henderson V. Assessing risk/benefit for trials using preclinical evidence: a proposal. J Med Ethics. 2016 Jan;42(1):50–3.
18. Kimmelman J, Federico C. Consider drug efficacy before first-in-human trials. Nature. 2017

Feb;542(7639):25–7.

19. Groeneveld GJ, Hay JL, Van Gerven JM. Measuring blood–brain barrier penetration using the NeuroCart, a CNS test battery. Drug Discovery Today: Technologies. 2016 Jun;20:27–34.

20. Teitelbaum Z, Lave T, Freijer J, Cohen AF. Risk Assessment in Extrapolation of Pharmacokinetics from Preclinical Data to Humans: Clinical Pharmacokinetics. 2010 Sep;49(9):619–32.

21. Peters SA, Petersson C, Blaukat A, Halle J-P, Dolgos H. Prediction of active human dose: learnings from 20 years of Merck KGaA experience, illustrated by case studies. Drug Discovery Today. 2020 Jan;S1359644620300283.

22. Morgan P, Van Der Graaf PH, Arrowsmith J, Feltner DE, Drummond KS, Wegner CD, et al. Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. Drug Discovery Today. 2012 May;17(9–10):419–24.

23. Mayawala K, de Alwis DP, Sachs JR. Model Informed Drug Development: Novel Oncology Agents are Lost in Translation. Clin Cancer Res. 2019 Nov 15;25(22):6564–6.

24. Blanchard OL, Smoliga JM. Translating dosages from animal models to human clinical trials—revisiting body surface area scaling. FASEB j. 2015 May;29(5):1629–34.

25. Davies M, Jones RDO, Grime K, Jansson-Löfmark R, Fretland AJ, Winiwarter S, et al. Improving the Accuracy of Predicted Human Pharmacokinetics: Lessons Learned from the AstraZeneca Drug Pipeline Over Two Decades. Trends in Pharmacological Sciences. 2020 Apr;S0165614720300687.

The Investigator's Brochure (IB) and the translational relevance of pharmacokinetic parameters

1) Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands;
2) Copernicus Institute of Sustainable Development, Innovation Studies, Utrecht University, Utrecht, The Netherlands;
3) Medicines Evaluation Board, Utrecht, The Netherlands.

6

Guilherme S. Ferreira[1],
Désirée H. Veening-Griffioen[1],
Wouter P.C. Boon[2],
Ellen H.M. Moors[2],
Peter J.K. van Meer[1,3]
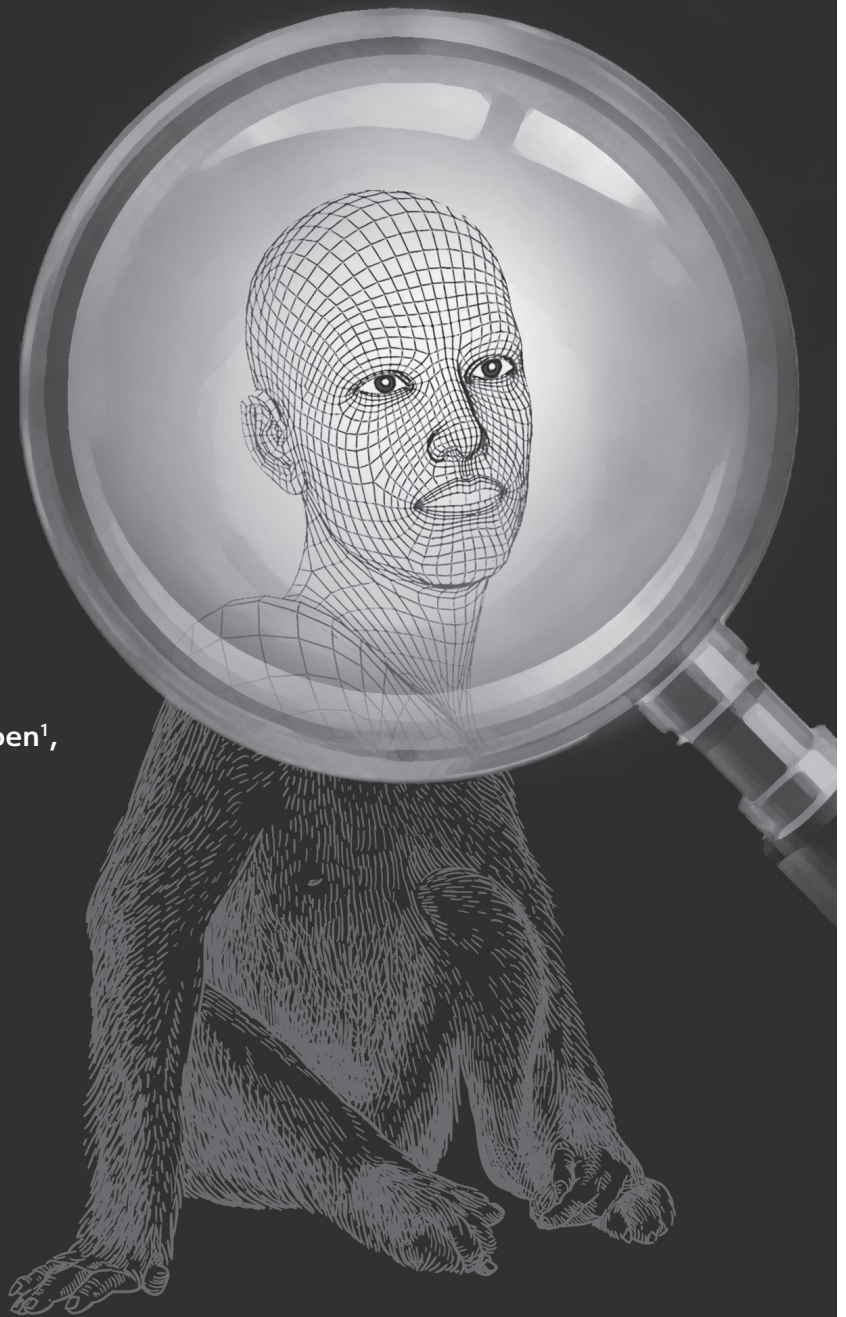
LEVELLING THE TRANSLATIONAL GAP FOR ANIMAL TO HUMAN EFFICACY DATA

Abstract

Reports of a reproducibility crisis combined with a high attrition rate in the pharmaceutical industry have put animal research increasingly under scrutiny in the past decade. Many researchers and the general public now question whether there is still a justification for conducting animal studies. While criticism of the current modus operandi in preclinical research is certainly warranted, the data on which these discussions are based are often unreliable. Several initiatives to address the internal validity and reporting quality of animal studies (e.g., Animals in Research: Reporting In Vivo Experiments (ARRIVE) and Planning Research and Experimental Procedures on Animals: Recommendations for Excellence (PREPARE) guidelines) have been introduced but seldom implemented. As for external validity, progress has been virtually absent. Nonetheless, the selection of optimal animal models of disease may prevent the conduct of clinical trials based on unreliable preclinical data. Here, we discuss three contributions to tackle the evaluation of the predictive value of animal models of disease themselves. First, we developed the Framework to Identify Models of Disease (FIMD), the first step to standardise the assessment, validation and comparison of disease models. FIMD allows the identification of which aspects of the human disease are replicated in the animals, facilitating the selection of disease models more likely to predict human response. Second, we show an example of how systematic reviews and meta-analyses can provide another strategy to discriminate between disease models quantitatively. Third, we explore whether external validity is a factor in animal model selection in the Investigator's Brochure (IB), and we use the IB-derisk tool to integrate preclinical pharmacokinetic and pharmacodynamic data in early clinical development. Through these contributions, we show how we can address external validity to evaluate the translatability and scientific value of animal models in drug development. However, while these methods have potential, it is the extent of their adoption by the scientific community that will define their impact. By promoting and adopting high-quality study design and reporting as well as a thorough assessment of the translatability of drug efficacy of animal models of disease, we will have robust data to challenge and improve the current animal research paradigm.

Keywords: animal model, drug development, translational research, FIMD, validation, systematic review, meta-analysis, investigator's brochure, external validity

Introduction

Despite modest improvements, the attrition rate in the pharmaceutical industry remains high (1–3). Although the explanation for such a low success is multifactorial, the lack of translatability of animal research has been touted as a critical aspect (4,5). Evidence of the problems of animal research's modus operandi has been mounting for almost a decade. Research has shown preclinical studies have major design flaws (e.g. low power, irrelevant endpoints), are poorly reported, or both – leading to unreliable data which ultimately means that animals are subjected to unnecessary suffering and clinical trial participants are potentially placed at risk (6,7). Increasing reports of failure to reproduce preclinical studies across several fields pointed to the need to increase current standards (8–10). In these terms, the assessment of two essential properties of translational research, internal and external validity, is jeopardised (11).

Internal validity refers to whether the findings of an experiment in defined conditions are true (4). Measures related to internal validity, such as randomisation and blinding, reduce or prevent several types of bias, and have a profound effect on study outcomes (12,13). Initiatives, such as the Animals in Research: Reporting In Vivo Experiments (ARRIVE) and Planning Research and Experimental Procedures on Animals: Recommendations for Excellence (PREPARE) guidelines and harmonized animal research reporting principles (HARRP), address the issues related to the poor design and reporting of animal experiments (14–16). While all of these initiatives can resolve most, if not all, issues surrounding internal validity, they are poorly implemented, and their uptake by all stakeholders is remarkably slow(13,17).

If progress on internal validity front has been insufficient, for external validity, it has been virtually absent. External validity is related to whether an experiment's findings can be extrapolated to other circumstances (e.g. animal to human translation). For external validity, while the study design plays a significant role (e.g. relevant endpoints and time to treatment), there is another often overlooked dimension: the animal models themselves (18). The pitfalls of using animals to simulate human conditions, such as different aetiology and lack of genetic heterogeneity, have been widely recognised for a long time (1,10,19,20). Nonetheless, the few efforts to address external validity to the same extent as internal validity are still insufficient (21,22).

The results of a sizeable portion of animal studies are unreliable (8,9,23). If we cannot fully trust the data generated by animal experiments, how can we assess their value? We argue that for the sensible evaluation of animal models of disease, we need to generate robust data first. To generate robust data, we need models

that simulate the human disease features to the extent that allows the reliable translation between species. Through the selection of optimal animal models of disease, we can potentially prevent clinical trials from advancing based on data unlikely to translate. Our research focuses on the design of methods and tools to this end. In the following sections, we discuss the development, applications, limitations and implications of the Framework to Identify Models of Disease (FIMD), systematic reviews and meta-analysis and the IB-derisk to evaluate the external validity of animal models of disease.

The Framework to Identify Models of Disease (FIMD)

The evaluation of preclinical efficacy often employs animal models of disease. Here, we use 'animal model of disease' for any animal model that simulates a human condition (or symptom) for which a drug can be developed, including testing paradigms. While safety studies are tightly regulated, including standardisation of species and study designs, there is hardly any guidance available for efficacy (24). New drugs often have new and unknown mechanisms of action, which require tailor-made approaches for their elucidation (24,25). As such, it would be counterproductive for regulatory agencies and companies alike to predetermine models or designs for efficacy as it is done for safety. However, in practice, this lack of guidance has contributed to the performance of studies with considerable methodological flaws (6,26).

The assessment of the external validity of animal models has traditionally relied on the criteria of face, construct and predictive validity (27,28). These concepts are generic and highly prone to user interpretation, leading to the analysis of disease models according to different disease parameters. This situation complicates an objective comparison between animal models. Newer approaches, such as the tool developed by Sams-Dodd and Denayer et al., can be applied to *in vitro* and *in vivo* models and consist of the simple scoring of five categories (species, disease simulation, face validity, complexity and predictivity) according to their proximity to the human condition (20,29). Nevertheless, they still fail to capture relevant characteristics involved in the pathophysiology and drug response, such as histology and biomarkers.

To address the lack of standardisation and the necessity of a multi-dimensional appraisal of animal models of disease, we developed the Framework to Identify Models of Disease (FIMD) (21,22). The first step in the development of FIMD was the identification of the core parameters used to validate disease models in the literature. Eight domains were identified: Epidemiology, Symptomatology and Natural History

(SNH), Genetics, Biochemistry, Aetiology, Histology, Pharmacology and Endpoints. More than 60% of the papers included in our scoping review used three or fewer domains. As it stood, the validation of animal models followed the tendency of the field as a whole: no standardisation nor integration of the characteristics frequently mentioned as relevant.

Based on these results, we drafted questions about each domain to determine the similarity to the human condition (Table 1). The sheet containing the answers to these questions and references thereof is called a validation sheet. The weighting and scoring system weights all domains equally. The final score can be visualised in a radar plot of the eight domains and, together with the validation sheet, facilitates the comparison of disease models at a high level. An example of radar plot is presented in Figure 1. However, we know that the domains can be of different importance depending on the disease. For example, in a genetic disease, such as Duchenne Muscular Dystrophy (DMD), the genetic domain has a higher weight than in type 2 diabetes, in which environmental factors (e.g. diet) play a significant role.

At first, we also designed a weighting and scoring system based on the intended indication to compare models. This system divided diseases into three groups according to their aetiology: genetic, external and multifactorial. By establishing a hierarchical relationship between the domains according to a disease's characteristics, we expected to allow a more sensitive scoring of the models. However, after reflection, the indication-based scoring was not included for two reasons. The first reason was the added complexity of this feature, which would require an additional setup to be executed in a framework that is already information-dense, possibly limiting its implementation. The second reason was the lack of validation of said feature since the numerical values were determined with a biological rationale but no mathematical/statistical basis. As such, any gain in score sensitivity would not necessarily indicate a difference between animal models. We settled for a generic system that weighted all domains equally until more data using this methodology are available. An example of radar plot is presented in Figure 1.

To account for the low internal validity of animal research, we added a reporting quality and risk of bias assessment for the pharmacological validation section. This section includes all studies in which a drug intervention was tested. The reporting quality parameters were based on the ARRIVE guidelines, and the risk of bias questions were extracted from the tool published by the SYstematic Review Center for Laboratory animal Experimentation (SYRCLE) (14,30). With this information, researchers can put pharmacological studies into context and evaluate how reliable the results are likely to be.

The final contribution of FIMD was a validation definition for animal models of disease. We grounded the definition of validation on the evidence provided for a

model's context of use, grading it into four levels of confidence. With this definition, we intentionally decoupled the connotation of a validated model being a predictive model. Rather, we reinforce that a validated animal model is a model with well-defined, reproducible characteristics.
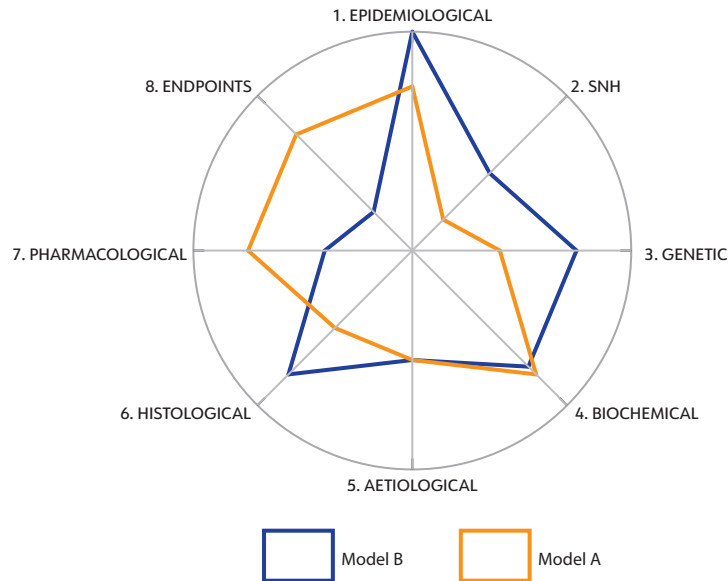

FIGURE 1 | Example of a radar plot obtained with the validation of two animal models using the Framework to Identify Models of Disease (FIMD. SNH - Symptomatology and Natural History. Extracted from Ferreira et al (21,22).

To validate our framework, we first conducted a pilot study of two models of type 2 diabetes – the Zucker Diabetic Fatty (ZDF) rat and db/db mouse, chosen on the basis of their extensive use in preclinical studies. Next, we did a complete validation of two models of Duchenne Muscular Dystrophy (DMD) – the mdx mouse and the Golden Retriever Muscular Dystrophy (GRMD) dog. We chose the mdx mouse owing to its common use as DMD model and the GRMD dog for its similarities to the human condition (31,32). While only minor differences were found for the type 2 diabetes models, the models for DMD presented more striking dissimilarities. The GRMD dog scored higher in the Epidemiology, SNH and Histology domains whereas the mdx mouse did so in the Pharmacology and Endpoints domains, the latter mainly driven by the absence of studies in the dog. Our findings indicate the mdx mouse may not be appropriate to test disease-modifying drugs, despite its use in most animal studies in DMD [31]. If more pharmacology studies are published using the GRMD dog, it will result into a more refined assessment. A common finding in all the four models was the high prevalence of experiments for which the risk of bias could not be assessed.

TABLE 1 | Questions per domain with weighting per questions. Extracted from Ferreira et al (21,22).

	Weight
1. EPIDEMIOLOGICAL VALIDATION	12.5
1.1 Is the model able to simulate the disease in the relevant sexes?	6.25
1.2 Is the model able to simulate the disease in the relevant age groups (e.g. juvenile, adult or ageing)?	6.25
2. SYMPTOMATOLOGY AND NATURAL HISTORY VALIDATION	12.5
2.1 Is the model able to replicate the symptoms and co-morbidities commonly present in this disease? If so, which ones?	2.5
2.2 Is the natural history of the disease similar to human's regarding: 2.2.1 Time to onset	2.5
2.2.2 Disease progression	2.5
2.2.3 Duration of symptoms	2.5
2.2.4 Severity	2.5
3. GENETIC VALIDATION	12.5
3.1 Does this species also have orthologous genes and/or proteins involved in the human disease?	4.17
3.2 If so, are the relevant genetic mutations or alterations also present in the orthologous genes/proteins?	4.17
3.3 If so, is the expression of such orthologous genes and/or proteins similar to the human condition?	4.16
4. BIOCHEMICAL VALIDATION	12.5
4.1 If there are known pharmacodynamic (PD) biomarkers related to the pathophysiology of the disease, are they also present in the model?	3.125
4.2 Do these PD biomarkers behave similarly to humans'?	3.125
4.3 If there are known prognostic biomarkers related to the pathophysiology of the disease, are they also present in the model?	3.125
4.4 Do these prognostic biomarkers behave similarly to humans'?	3.125
5. AETIOLOGICAL VALIDATION	12.5
5.1 Is the aetiology of the disease similar to humans'?	12.5
6. HISTOLOGICAL VALIDATION	12.5
6.1 Do the histopathological structures in relevant tissues resemble the ones found in humans?	12.5
7. PHARMACOLOGICAL VALIDATION	12.5
7.1 Are effective drugs in humans also effective in this model?	4.17
7.2 Are ineffective drugs in humans also ineffective in this model?	4.17
7.3 Have drugs with different mechanisms of action and acting on different pathways been tested in this model? If so, which?	4.16
8. ENDPOINT VALIDATION	12.5
8.1 Are the endpoints used in preclinical studies the same or translatable to the clinical endpoints?	6.25
8.2 Are the methods used to assess preclinical endpoints comparable to the ones used to assess related clinical endpoints?	6.25

We designed FIMD to avoid the limitations of previous approaches, which included the lack of standardisation and integration of internal and external validities. Nonetheless, it presents challenges of its own, ranging from the definition of disease parameters and absence of a statistical model to support a more sensitive weighting and scoring system to the use of publicly available (and often biased) data. The latter is especially relevant, as study design and reporting deficiencies of animal studies undoubtedly represent a challenge to interpret the resulting data correctly. While owing to these deficiencies, many studies are less informative; some may still offer potential insights if the data are interpreted adequately. We included the reporting quality and risk of bias assessment to force researchers to account for these shortcomings when interpreting the data.

FIMD integrates the key aspects of the human disease an animal model must simulate. Naturally, no model is expected to mimic the human condition fully. However, understanding which features an animal model can and which it cannot replicate, allows researchers to select optimal models for their research question. More importantly, it puts the results from animal studies into the broader context of human biology, potentially preventing the advancement of clinical trials based on data unlikely to translate.

Systematic Review and Meta-analysis

Systematic reviews and meta-analyses of animal studies were one of the earliest tools to expose the status of the field (33). Nonetheless, their application to compare animal models of disease is relatively recent. In FIMD's pilot study, the two models of type 2 diabetes (the ZDF rat and db/db mouse) only presented slight differences in the general score. Since FIMD does not compare models quantitatively, we conducted a systematic review and meta-analysis to compare the effect of glucose-lowering agents approved for human use on the HbA1c (34). We chose HbA1c as the outcome owing to its clinical relevance in type 2 diabetes drug development (35,36).

The results largely confirmed FIMD's pilot study results: both models responded similarly to drugs irrespective of the mechanism of action. The only exception was exenatide, which led to higher reductions of HbA1c in the ZDF rat. Both models predicted the direction of effect in humans for drugs with enough studies. Moreover, the quality assessment showed that animal studies are poorly reported: no study mentioned blinding at any level, and less than half reported randomisation. In this context, the risk of bias could not be reliably assessed.

The development of systematic reviews and meta-analyses to combine animal and human data offers an unprecedented opportunity to investigate the value of animal models of disease further. Nevertheless, translational meta-analyses are still uncommon (37). A prospect for another application of systematic reviews and meta-analyses lies in comparing drug effect sizes in animals and humans directly. By calculating the degree of overlap of preclinical and clinical data, animal models could be ranked according to the extent they can predict effect sizes across different mechanisms of action and drug classes. The calculation of this 'translational coefficient' would include effective and ineffective drugs. Using human effect sizes as the denominator, a translational coefficient higher than 1 would indicate an overestimation of treatment effect while a coefficient lower than 1, an underestimation. The systematisation of translational coefficients would lead to 'translational tables', giving additional insight on models' translatability. These translational tables, allied with more qualitative approaches, such as FIMD, could form the basis for evidence-based animal model selection in the future.

Indeed, such a strategy is not without shortcomings. Owing to significant differences in the methodology of preclinical and clinical studies, such comparisons may present unusually large confidence intervals, complicating their interpretation. Also, preclinical studies would need to match the standards of clinical research to a higher degree, including the use of more relevant endpoints that can be compared. Considerations on other design (e.g. dosing, route of administration), statistical (e.g. sample size, measures of spread) and biological matters (species differences) will be essential to develop a scientifically sound approach.

Can animal models predict human pharmacologically active ranges? A first glance into the Investigator's Brochure

The decision to proceed to first-in-human trials is mostly based on the Investigator's Brochure (IB), a document required by the Good Clinical Practice (GCP) guidelines (38). The IB compiles all the necessary preclinical and clinical information for ethics committees and investigators to evaluate a drug's suitability to be tested in humans. The preclinical efficacy and safety data in the IB are thus the basis for the risk-benefit analysis at this stage. Therefore, it is paramount that these experiments are performed to the highest standards to safeguard healthy volunteers and patients.

However, the results from Wieschoswki and colleagues show a different scenario (26). They analysed 109 IBs presented for ethics review of three German institutional review boards. The results showed that the vast majority of preclinical efficacy

studies did not report measures to prevent bias, such as blinding or randomisation. Furthermore, these preclinical studies were hardly ever published in peer-reviewed journals and were overwhelmingly positive – only 6% of the studies reported no significant effect. The authors concluded IBs do not provide enough high-quality data to allow a proper risk-benefit evaluation of investigational products for first-in-human studies during the ethics review.

In an ongoing study to investigate the predictivity of the preclinical data provided in the IBs, we are evaluating whether animal models can predict pharmacologically active ranges in humans. Since pharmacokinetic (PK) and pharmacodynamic (PD) data are often scattered throughout the IB across several species, doses and experiments, integrating it can be challenging. We are using the IB-derisk, a tool developed by the Centre for Human Drug Research (CHDR), to facilitate this analysis (39). The IB-derisk consists of a colour-coded excel sheet or web application (www.ib-derisk.org) in which PK and PD data can be inputted. It allows the extra- and interpolation of missing PK parameters across animal experiments, facilitating the dose selection in first-in-human trials. The IB-derisk yields yet another method to discriminate between animal models of disease. With sufficient data, the drug PK and PD from preclinical studies of a single model and clinical trials of correspondent drugs can be compared. This analysis, when combined with PK/PD modelling, can serve as a tool to select the most relevant animal model based on the mechanism of action and model characteristics. Preliminary (and unpublished) results suggest that animal models can often predict human pharmacologically active ranges. How the investigated pharmacokinetic parameters relate to the indication, safety, and efficacy is still unclear.

To build on Wieschowski's results, we have been collecting data on the internal validity and reporting quality of animal experiments. Our initial analysis indicates the included IBs also suffer from the same pitfalls identified by Wieschowski, suggesting that such problems are likely to be widespread. Also, only a few IBs justified their choice of model(s) of disease, and none compared their model(s) to other options to better understand their pros and cons. This missing information is crucial to allow risk-benefit analysis during the ethics review process.

Levelling the translational gap for animal to human efficacy data

Together, FIMD, systematic reviews and meta-analyses and the IB-derisk have the potential to circumvent many limitations of current preclinical efficacy research. FIMD allows the general characterisation of animal models of disease, enabling the identification of their strengths and weaknesses. The validation process results

in a sheet and radar plot that can be compared easily. Because the validation is indication-specific, the same model may have different scores for different diseases, allowing a more nuanced assessment. The addition of measures of the risk of bias and reporting quality guarantees researchers have enough information to scrutinise the efficacy data. The IB-derisk tool correlates human and animal PK and PD ranges, adding a quantitative measure of response to the pharmacological validation of FIMD. Finally, a systematic review and meta-analysis of efficacy studies can show whether the pharmacological effects in animals translate into actual clinical effects.

This extensive characterisation of disease models – and their disparities to patients, is the only way to account for species differences that may impact the translation (11,40). The largest study so far to analyse the concordance between animal and human safety data supports this premise (41). For example, while rabbits can accurately predict arrhythmia in humans, the Irwin test in rats, a required safety pharmacology assessment, is of limited value. Thus, we need to focus on understanding the animal pathophysiology to the degree that allows us to assess for what kind of research a disease model is suitable. Especially for complex multifactorial conditions (e.g. Alzheimer's disease), the use of multiple models that simulate different aspects is likely to provide a more detailed and reliable picture (42). For instance, Seok and colleagues have shown that human and mouse responses to inflammation vary significantly according to their aetiology (43). While genomic responses in human trauma and burn were highly correlated ($R^2 = 0.91$), human to mouse responses correlated poorly with an $R^2 < 0.1$. These considerations are fundamental to select a model of disease to assess the efficacy of potential treatments, and FIMD can serve as the tool to compile and integrate this information. However, by itself, FIMD cannot prevent poor models from being used. The validation of all existing models may result in low scores, and thus, the inability to identify a relevant model. While the researcher may still choose to pick one of these models for a specific reason, FIMD offers institutional review boards and funders a scientifically-sound rationale to refuse the performance of studies in poor animal models – even if it means no animal research should be conducted.

There are some constraints for the implementation of the methods and tools we described. All of them require significant human and financial investments. Nonetheless, the estimated cost of irreproducibility (~US$30 billion per year in the US alone) – likely dwarfs the cost for the broad implementation of these strategies (44). Besides, this initial investment can be mitigated over time. For instance, FIMD validation sheets could be available in peer-reviewed, open-access publications that can be updated periodically. This availability will prevent researchers from different institutions from unnecessarily repeating studies. Also, subsequent FIMD updates will be much faster than the first validation. In the same vein, if the application of

Levelling the translational gap for animal to human efficacy data

systematic reviews and meta-analyses and the IB-derisk become commonplace, training will become widespread, and their execution will be more efficient. The development of automated methods that can compile vast amounts of data will certainly aid to this end (45,46).

Animal research is already a cost-intensive and often long endeavour. By conducting experiments with questionable validities, we are misusing animals – which is expressively prohibited by the European Union (EU) Directive 2010/63. Only with a joint effort involving researchers in academia and industry, ethics committees, funders, regulatory agencies and the pharmaceutical industry, we can improve the quality of animal research. By applying FIMD, systematic reviews and meta-analysis, and the IB-derisk, researchers can identify more predictive disease models, potentially preventing clinical trials starting based on unreliable data. These approaches can be implemented in the short-term in both the academic and industrial settings since the training requires only a few months.

Concomitantly, the other stakeholders must create an environment that encourages the adoption of best practices. Ethics committees have a unique opportunity to incentivise higher standards since an unfavourable assessment can prevent poorly designed experiments from even starting. However, they now frequently base their decisions on subpar efficacy data (26). Also, the lack of a detailed disease model justification often results in the selection of disease models based on tradition, rather than science (47). The request of a more detailed translational rationale for each model choice (e.g. by requiring models are evaluated with FIMD), as well as the enforcement of reporting guidelines, can act as gatekeepers for flawed study designs and improve the risk-benefit analysis significantly (48).

Funders can require the use of systematic reviews and meta-analyses and a thorough assessment of the translational relevance of selected animal models (e.g. FIMD). They can facilitate the adoption of these measures by reserving some budget specifically for training and the implementation of these strategies. Over time, FIMD, systematic reviews and meta-analyses can eventually become an essential requirement to acquire funding. Provided there is a grace period of at least a year, funders could request a stricter justification of model selection as early as their next grant round.

Journal editors and reviewers must actively enforce reporting guidelines for ongoing submissions as endorsing them does not improve compliance (49). Promoting adherence to higher reporting standards will also preserve the journal's reputation. In the medium-term, journals can provide a list with relevant parameters (e.g. randomisation, blinding) with line references to be filled out by the authors in the submission system. Such a system would facilitate reporting quality checks by editors

and reviewers without increasing the review length substantially. Registered reports allow input on the study design since they are submitted before the beginning of experiments. Additionally, since acceptance is granted before the results are known, the registered reports also encourage the publication of negative results, acting against the publication bias for positive outcomes.

Regulatory agencies can shape the drug development landscape significantly with some key actions. For instance, updating the IB guidelines by requiring a more extensive translational rationale for each animal model employed would facilitate the risk-benefit analysis by assessors and ethics committees alike. This rationale should include not only an evaluation of the animal model itself but also how it compares to other available options. Furthermore, agencies could review disease-specific guidance to include a more comprehensive account of efficacy assessment by exploring the use of disease models in safety studies (24,50). The simultaneous evaluation of efficacy and safety can result in more informative studies, which are more likely to translate to the clinic. Finally, the output of the translational assessments of animal models of disease can be incorporated within periodic updates of guidelines, for instance, as an extended version of Sheean and colleagues' work (51). Scientific advice can be used as a platform to discuss translational considerations early in development. By presenting the strengths and weaknesses of validated disease models, agencies can promote optimal model selection without precluding model optimisation and the development of new approaches.

Furthermore, drug development companies can significantly benefit from the implementation of these measures in the medium-term. Larger companies can perform a thorough assessment of preclinical data of internal and external assets using FIMD, systematic review and meta-analysis, and the IB-derisk. At the same time, small and medium enterprises can provide data in these formats to support their development plan. Ultimately, the selection of more predictive disease models will lead to more successful clinical trials, increasing the benefit and reducing the risks to patients, and lower development costs.

A positive side-effect of these strategies is the increased scrutiny of design and animal model choices. Instead of a status-quo based on tradition and replication of poor practices, we can move forward to an inquisitive and evidence-based modus operandi (7). This change of culture is sorely needed in both academic and industrial institutions (52). A shift toward a stricter approach – more similar to clinical trials, from beginning to end, is warranted (52).

This shift should include the, possibly immediate, preregistration and publication of animal studies on a public online platform, such as preclinicaltrials.eu (6,7,53–55). As

with clinical trials, where the quality of the preregistration and the actual publication of results are still a concern, this is unlikely to be sufficient (56,57). Higher standards of study design, performance and reporting are necessary to increase the quality of the experiments. Currently, Good Laboratory Practices (GLP) are only required for preclinical safety but not efficacy assessment. However, GLP experiments can be costly, making them prohibitive for academia, where many animal studies are conducted (58). A change toward the Good Research Practices (GRP), published by the World Health Organisation, offers a feasible alternative (59). The GRP consists of several procedures to improve the robustness of data and reproducibility, similarly to the GLP, but less strict. With the combination of efficacy and safety studies in disease models, the application of GRP instead of GLP may also be implemented in academia. Together with the employment of internal (e.g. ARRIVE and PREPARE guidelines, HARRP) and external validity (e.g. FIMD, systematic review and meta-analysis, IB-derisk) tools, these proposals tackle critical shortcomings of current animal research. The harmonisation of requirements across stakeholders will be crucial for a successful change of mindset.

Our suggestions to improve the preliminary assessment of efficacy in animals are likely to face resistance from a considerable part of the scientific community (60). Nevertheless, so did the introduction of actions to improve the quality of clinical trials (53). The requirement for preregistration and the establishment of blinded and randomised clinical trials as the gold standard have improved clinical research substantially. The current requirements for clinical trials to be executed are also time-consuming, but few would doubt their importance. It is time we apply the same rigour to animal research (48,54,61).

At present, the discussions over the translatability of animal studies are often based on questionable data. If many preclinical studies are poorly conducted, we cannot expect their results to be translatable to the clinical context. Another overlooked point is the definition of translatability standards. Provided that studies are perfectly designed, employ optimal models and are conducted meticulously, what is the level of concordance that society and scientists are willing to accept? This lack of standardisation leads to the extraordinary situation of both proponents and opponents of the current animal research paradigm citing the same data to defend their arguments (41,62,63).

This reflection is especially relevant for the discussion of alternatives to animal studies. Many *in silico* and *in vitro* (particularly organoids and organs-on-a-chip) approaches are in development aiming to replace animal use in drug development partially or entirely, but they also have limitations (50,64). For instance, organoids cannot simulate organ-organ interactions, and the organ-on-a-chip technology has

not resolved the lack of a universal media to connect different organs as well as its inability to replicate the immune response, endocrine system or gut-microbiome (65). If these systems have the edge over animal studies owing to their human origin, they do not reproduce the known and unknown interactions of a whole intact organism. Additionally, they encounter challenges related to validation, reproducibility and reporting, akin to their animal counterparts (66–68). Hence, in their present form, these approaches are better applied as a complement, rather replacement, to animal research.

In the foreseeable future, animal research is unlikely to be eliminated. The criticism of the present state of affairs of preclinical research is indeed justifiable. However, dismissing the value of animals on the grounds of questionable data seems excessive. Meanwhile, our efforts must be focused on improving the robustness of animal data generated now. We already have tools available to address most, if not all, internal and external validity concerns. Only a thorough assessment of higher-quality animal data will determine whether animal research is still a valid paradigm in drug development.

Final considerations

As it stands, the scepticism over the justification of animal research is well-founded. The data on which we routinely base our decisions in drug development is often unreliable. A significant reappraisal of the current standards for animal research is warranted. We must design, conduct and report preclinical studies with the same rigour as clinical studies. Moreover, we must scrutinise animal models of disease to ensure they are relevant to the evaluation of preliminary efficacy or any other question we are asking. If there are no relevant animal models of disease, then preclinical testing in other platforms must be pursued.

Those changes will undoubtedly have a cost. Nevertheless, all stakeholders must be willing to invest human and financial resources to drive a change in culture and practice. Only when we promote and adopt a high-quality study design and reporting as well as a thorough assessment of animal models' translatability, we will have robust data to challenge and improve the current paradigm.

Acknowledgments

Levelling the translational gap for animal to human efficacy data

Conflicts of Interest

GSF reports personal fees from Merck KGaA and Curare Consulting B.V. outside of the submitted work. DVG reports personal fees from Nutricia Research B.V and Merck & Co outside of the submitted work. None of the other authors has any conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov. 2004;3(8):711–5.
2. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019 Apr 1;20(2):273–86.
3. Pammolli F, Righetto L, Abrignani S, Pani L, Pelicci PG, Rabosio E. The endless frontier? The recent increase of R&D productivity in pharmaceuticals. J Transl Med. 2020 Dec;18(1):162.
4. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can Animal Models of Disease Reliably Inform Human Studies? PLoS Med. 2010 Mar 30;7(3):e1000245.
5. Schulz JB, Cookson MR, Hausmann L. The impact of fraudulent and irreproducible data to the translational research crisis - solutions and implementation. J Neurochem. 2016;139 Suppl 2:253–70.
6. Ioannidis JPA. Acknowledging and Overcoming Nonreproducibility in Basic and Preclinical Research. JAMA. 2017 Mar 14;317(10):1019.
7. Vogt L, Reichlin TS, Nathues C, Würbel H. Authorization of Animal Experiments Is Based on Confidence Rather than Evidence of Scientific Rigor. Dirnagl U, editor. PLoS Biol. 2016 Dec 2;14(12):e2000598.
8. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012 Mar 28;483(7391):531–3.
9. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov. 2011 Sep;10(9):712–712.
10. Perrin S. Preclinical research: Make mouse studies work. Nature. 2014 Mar;507(7493):423–5.
11. Pound P, Ritskes-Hoitinga M. Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. J Transl Med. 2018 Dec;16(1):304.
12. Bebarta V, Luyten D, Heard K. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad Emerg Med. 2003 Jun;10(6):684–7.
13. Schmidt-Pogoda A, Bonberg N, Koecke MHM, Strecker J, Wellmann J, Bruckmann N, et al. Why Most Acute Stroke Studies Are Positive in Animals but Not in Patients: A Systematic Comparison of Preclinical, Early Phase, and Phase 3 Clinical Trials of Neuroprotective Agents. Ann Neurol. 2020 Jan;87(1):40–51.
14. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLoS Biol. 2010 Jun 29;8(6):e1000412.
15. Smith AJ, Clutton RE, Lilley E, Hansen KEA, Brattelid T. PREPARE: guidelines for planning animal research and testing. Lab Anim. 2018 Apr;52(2):135–41.
16. Osborne N, Avey MT, Anestidou L, Ritskes-Hoitinga M, Griffin G. Improving animal research reporting standards: HARRP, the first step of a unified approach by ICLAS to improve animal research reporting standards worldwide. EMBO Rep [Internet]. 2018 May [cited 2020 May 19];19(5). Available from: https://onlinelibrary.wiley.com/doi/abs/10.15252/embr.201846069
17. Baker D, Lidster K, Sottomayor A, Amor S. Two Years Later: Journals Are Not Yet Enforcing the ARRIVE Guidelines on Reporting Standards for Pre-Clinical Animal Studies. Eisen JA, editor. PLoS Biol. 2014 Jan 7;12(1):e1001756.
18. Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG. Threats to Validity in the Design and Conduct of Preclinical Efficacy Studies: A Systematic Review of Guidelines for In Vivo Animal Experiments. Ioannidis JPA, editor. PLoS Med. 2013 Jul 23;10(7):e1001489.
19. Hackam DG. Translating animal research into clinical benefit. BMJ. 2007 Jan 27;334(7586):163–4.
20. Sams-Dodd F. Strategies to optimize the validity of disease models in the drug discovery process. Drug Discovery Today. 2006 Apr;11(7–8):355–63.

21. S. Ferreira G, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, Schellekens H, et al. A standardised framework to identify optimal animal models for efficacy assessment in drug development. Macleod MR, editor. PLoS ONE. 2019 Jun 13;14(6):e0218014.

22. Ferreira GS, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, Schellekens H, et al. Correction: A standardised framework to identify optimal animal models for efficacy assessment in drug development. PLoS ONE. 2019 Jul 22;14(7):e0220325.

23. Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. PLoS Biol. 2015 Oct 13;13(10):e1002273.

24. Langhof H, Chin WWL, Wieschowski S, Federico C, Kimmelman J, Strech D. Preclinical efficacy in therapeutic area guidelines from the U.S. Food and Drug Administration and the European Medicines Agency: a cross-sectional study. Br J Pharmacol. 2018;175(22):4229–38.

25. Varga OE, Hansen AK, Sandøe P, Olsson IAS. Validating animal models for preclinical research: a scientific and ethical discussion. Altern Lab Anim. 2010 Jun;38(3):245–8.

26. Wieschowski S, Chin WWL, Federico C, Sievers S, Kimmelman J, Strech D. Preclinical efficacy studies in investigator brochures: Do they enable risk–benefit assessment? Macleod M, editor. PLoS Biol. 2018 Apr 5;16(4):e2004879.

27. McKinney WT. Animal Model of Depression: I. Review of Evidence: Implications for Research. Arch Gen Psychiatry. 1969 Aug 1;21(2):240.

28. Willner P. The validity of animal models of depression. Psychopharmacology. 1984 Apr;83(1):1–16.

29. Denayer T, Stöhr T, Roy MV. Animal models in translational medicine: Validation and prediction. European Journal of Molecular & Clinical Medicine. 2014 Aug 27;2(1):5.

30. Hooijmans CR, Rovers MM, de Vries RBM, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol. 2014 Mar 26;14:43.

31. McGreevy JW, Hakim CH, McIntosh MA, Duan D. Animal models of Duchenne muscular dystrophy: from basic mechanisms to gene therapy. Dis Model Mech. 2015 Mar;8(3):195–213.

32. Yu X, Bao B, Echigoya Y, Yokota T. Dystrophin-deficient large animal models: translational research and exon skipping. Am J Transl Res. 2015;7(8):1314–31.

33. Pound P, Ebrahim S, Sandercock P, Bracken MB, Roberts I. Where is the evidence that animal research benefits humans? BMJ. 2004 Feb 28;328(7438):514–7.

34. Ferreira GS, Veening-Griffioen DH, Boon WPC, Hooijmans CR, Moors EHM, Schellekens H, et al. Comparison of drug efficacy in two animal models of type 2 diabetes: A systematic review and meta-analysis. European Journal of Pharmacology. 2020 Jul;879:173153.

35. FDA. Guidance for Industry Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention [Internet]. Food and Drug Administration; 2008 [cited 2019 Oct 14]. Available from: https://www.fda.gov/media/71289/download

36. EMA. Guideline on clinical investigation of medicinal products in 5 the treatment or prevention of diabetes mellitus [Internet]. European Medicines Agency; 2018 [cited 2019 Oct 14]. Available from: https://www.ema.europa.eu/documents/scientific-guideline/draft-guideline-clinical-investigation-medicinal-products-treatment-prevention-diabetes-mellitus_en.pdf

37. Leenaars CHC, Kouwenaar C, Stafleu FR, Bleich A, Ritskes-Hoitinga M, De Vries RBM, et al. Animal to human translation: a systematic scoping review of reported concordance rates. J Transl Med. 2019 Dec;17(1):223.

38. European Medicines Agency. ICH E6 (R2) Good clinical practice - Step 5 [Internet]. 2016 [cited 2020 Apr 28]. Available from: https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice

39. van Gerven J, Cohen A. Integrating data from the Investigational Medicinal Product Dossier/ investigator's brochure. A new tool for translational integration of preclinical effects. Br J Clin Pharmacol. 2018 Jul;84(7):1457–66.

40. Zeiss CJ. Improving the predictive value of interventional animal models data. Drug Discovery Today. 2015 Apr;20(4):475–82.

41. Clark M, Steger-Hartmann T. A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans. Regulatory Toxicology and Pharmacology. 2018 Jul;96:94–105.

42. Veening-Griffioen DH, Ferreira GS, van Meer PJK, Boon WPC, Gispen-de Wied CC, Moors EHM, et al. Are some animal models more equal than others? A case study on the translational value of animal models of efficacy for Alzheimer's disease. European Journal of Pharmacology. 2019 Sep;859:172524.

43. Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc Natl Acad Sci USA. 2013 Feb 26;110(9):3507–12.

44. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. PLoS Biol. 2015 Jun 9;13(6):e1002165.

45. Bannach-Brown A, Przybyła P, Thomas J, Rice ASC, Ananiadou S, Liao J, et al. Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. Syst Rev. 2019 Dec;8(1):23.

46. Zeiss CJ, Shin D, Vander Wyk B, Beck AP, Zatz N, Sneiderman CA, et al. Menagerie: A text-mining tool to support animal-human translation in neurodegeneration research. Forloni G, editor. PLoS ONE. 2019 Dec 17;14(12):e0226176.

47. Veening-Griffioen DH, Ferreira GS, Boon WPC, Gispen-de Wied CC, Schellekens H, Moors EHM, et al. Tradition, not science, is the basis of animal model selection in translational and applied research. ALTEX - Alternatives to animal experimentation.

48. Kimmelman J, Henderson V. Assessing risk/benefit for trials using preclinical evidence: a proposal. J Med Ethics. 2016 Jan;42(1):50–3.

49. Hair K, Macleod MR, Sena ES. A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). Res Integr Peer Rev. 2019 Dec;4(1):12.

50. Avila AM, Bebenek I, Bonzo JA, Bourcier T, Davis Bruno KL, Carlson DB, et al. An FDA/CDER perspective on nonclinical testing strategies: Classical toxicology approaches and new approach methodologies (NAMs). Regulatory Toxicology and Pharmacology. 2020 Jul;114:104662.

51. Sheean ME, Malikova E, Duarte D, Capovilla G, Fregonese L, Hofer MP, et al. Nonclinical data supporting orphan medicinal product designations in the area of rare infectious diseases. Drug Discovery Today. 2020 Feb;25(2):274–91.

52. Howells DW, Sena ES, Macleod MR. Bringing rigour to translational medicine. Nat Rev Neurol. 2014 Jan;10(1):37–43.

53. Begley CG, Ioannidis JPA. Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. Circ Res. 2015 Jan 2;116(1):116–26.

54. Kimmelman J, Anderson JA. Should preclinical studies be registered? Nat Biotechnol. 2012 Jun;30(6):488–9.

55. PreclinicalTrials.eu. PreclinicalTrials.eu [Internet]. [cited 2020 May 20]. Available from: https://preclinicaltrials.eu/

56. Viergever RF, Karam G, Reis A, Ghersi D. The Quality of Registration of Clinical Trials: Still a Problem. Scherer RW, editor. PLoS ONE. 2014 Jan 10;9(1):e84727.

57. DeVito NJ, Bacon S, Goldacre B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. The Lancet. 2020 Feb;395(10221):361–9.

58. van Meer PJK, Graham ML, Schuurman H-J. The safety, efficacy and regulatory triangle in drug development: Impact for animal models and the use of animals. European Journal of Pharmacology. 2015 Jul;759:3–13.

59. World Health Organization. Handbook: Quality practices in basic biomedical research [Internet]. 2006 [cited 2020 May 23]. Available from: https://www.who.int/tdr/publications/training-guideline-publications/handbook-quality-practices-biomedical-research/en/

Levelling the translational gap for animal to human efficacy data

60. ter Riet G, Korevaar DA, Leenaars M, Sterk PJ, Van Noorden CJF, Bouter LM, et al. Publication Bias in Laboratory Animal Research: A Survey on Magnitude, Drivers, Consequences and Potential Solutions. von Elm E, editor. PLoS ONE. 2012 Sep 5;7(9):e43404.

61. Kimmelman J, Federico C. Consider drug efficacy before first-in-human trials. Nature. 2017 Feb;542(7639):25–7.

62. Bailey J, Balls M. Recent efforts to elucidate the scientific validity of animal-based drug tests by the pharmaceutical industry, pro-testing lobby groups, and animal welfare organisations. BMC Med Ethics. 2019 Dec;20(1):16.

63. Monticello TM, Jones TW, Dambach DM, Potter DM, Bolt MW, Liu M, et al. Current nonclinical testing paradigm enables safe entry to First-In-Human clinical trials: The IQ consortium nonclinical to clinical translational database. Toxicology and Applied Pharmacology. 2017 Nov;334:100–9.

64. Van Norman GA. Limitations of Animal Studies for Predicting Toxicity in Clinical Trials. JACC: Basic to Translational Science. 2020 Apr;5(4):387–97.

65. Haddrick M, Simpson PB. Organ-on-a-chip technology: turning its potential for clinical benefit into reality. Drug Discovery Today. 2019 May;24(5):1217–23.

66. Vives J, Batlle-Morera L. The challenge of developing human 3D organoids into medicines. Stem Cell Res Ther. 2020 Dec;11(1):72.

67. Mead BE, Karp JM. All models are wrong, but some organoids may be useful. Genome Biol. 2019 Dec;20(1):66.

68. Hartung T, De Vries R, Hoffmann S, Hogberg HT, Smirnova L, Tsaioun K, et al. Toward Good In Vitro Reporting Standards. ALTEX. 2019;36(1):3–17.

Summary

Samenvatting

Resumo

Acknowledgements

Curriculum vitae

List of publications

Summary

Differently from the 1990s, when 40% of drugs failed in humans because of pharmacokinetics and bioavailability issues, now the primary reasons for failure are clinical efficacy, safety and commercial, in this order. Since new drugs frequently advance into the clinic based on animal studies, insufficient efficacy or safety in humans suggests that animal models are not predictive enough. Poor reproducibility has often been cited as a primary cause for this high failure rate. For this reason, most efforts so far focused on addressing the poor quality in the design, conduct and reporting of animal studies (internal validity). Currently, an improvement of the internal validity depends on the implementation, rather than development, of guidelines and recommendations.

Conversely, the evaluation of animal models themselves (external validity) has often been overlooked. The current ways to assess the external validity of animal models of disease have several limitations. They are not standardised, may be interpreted differently by each user, preclude an easy comparison between animal models or lacks a more in-depth analysis of pathophysiological features relevant for drug efficacy. Nonetheless, it is only by establishing the essential information through which animal models can be evaluated that we can select models more likely to predict the human response. Hence, there is a need for new ways to discriminate among potentially predictive animal models of disease.

In this thesis, we developed new ways to discriminate among potentially predictive animal models. We defined the standards to guide the assessment, comparison and validation of animal models used to test the preliminary efficacy of new drugs. We based our approach on outlining the relationships between animal and human disease at physiological, pharmacological and therapeutic levels.

In **Chapter 2**, we present the development of the Framework to Identify Models of Disease (FIMD). We identified in the literature eight domains used by researchers in the validation and optimisation of animal models of disease: Epidemiology, Symptomatology and Natural History (SNH), Genetics, Biochemistry, Aetiology, Histology, Pharmacology and Endpoints. Based on these domains, questions were drafted to assess the similarities between animal and human condition. We included a weighting and scoring system that weighted all domains equally. Moreover, FIMD features measures of comparability between animal models, uncertainty around the validation, reporting quality and risk of bias. FIMD's completed with instructions to facilitate its application in the assessment, validation and comparison of animal models of disease.

In **Chapter 3**, we detail the validation of FIMD, which included a simplified version of the framework (pilot study) for two animal models of Type 2 Diabetes (T2D) and a complete validation for two models of Duchenne Muscular Dystrophy (DMD). In the pilot study, our results of the T2D models Zucker Diabetic Fatty (ZDF) rat and db/db mouse suggested that there are only minor differences between them, although the Pharmacology domain lacked enough studies. Conversely, the complete validation of the DMD models mdx mouse and Golden Retriever Muscular Dystrophy (GRMD) dog, showed significant differences in five domains: Epidemiology, SNH, Histology, Pharmacology and Endpoints. While the GRMD dog scored better in the first three, the mdx mouse was superior in the last two. Nevertheless, the higher score of the mdx mouse may be partially attributed to the absence of studies in the dog, leading to a lower score in the Pharmacology and Endpoints domains. The reporting quality was poor, and the risk of bias could not be assessed in most cases. Using the complete validation of the DMD models as examples, we introduce a step-wise approach to perform a complete validation of any animal model of disease.

In **Chapter 4**, we further explored the results from FIMD's pilot study in T2D, which showed only minor differences between the ZDF rat and db/db mouse. Since FIMD did not fully investigate the concordance between animal and human data at therapeutic level, we conducted a systematic review and meta-analysis of animal studies to validate the pilot study results. The glucose-lowering effect of approved drugs on the glycated haemoglobin A1c (HbA1c) was compared in both models. Our findings indicated that the few dissimilarities between the ZDF rat and db/db mouse did not result in differences in the therapeutic effect of all drugs, except for exenatide. Similar to the pilot study results, because of the low quality of reporting, the risk of bias could not be evaluated for most studies. This systematic review and meta-analysis establishes the basis for the application of this methodology to compare animal and human efficacy data.

In **Chapter 5**, we explore preclinical pharmacokinetic and pharmacodynamic predictions in early clinical development to assess disease models at the pharmacological level. Investigator's Brochures (IBs), a critical document that summarises all relevant information before first-in-human studies, are used as a source material. We employed the IB-derisk tool (developed at the Centre for Human Drug Research – CHDR, Leiden, NL) to integrate animal and human pharmacokinetic and pharmacodynamic data. Our results suggest that while animal models can potentially predict human well-tolerated and pharmacologically active ranges, there is much room for improvement. The reporting quality of the IB was similar to our previous assessments in Chapters 3 and 4, indicating that the poor reporting is not restricted to scientific literature. We propose stricter requirements for the design, report and conduct of animal experiments, broader exploration of

exposure in efficacy studies, among other measures to improve the potential of the IB to aid animal to human translation.

Finally, in **Chapter 6**, the discussion, we recapitulate each method debated in this thesis: FIMD, systematic reviews and meta-analysis and the IB-derisk tool. We discuss their applications, challenges for implementation and impact on animal research. The use of these new methods to discriminate between animal models of disease can only go so forth in addressing the high attrition rates in drug development. Only with a joint effort involving researchers in academia and industry, ethics committees, funders, regulatory agencies, and the pharmaceutical industry, the quality of animal research can be improved.

By applying FIMD, systematic reviews and meta-analysis, and the IB-derisk, researchers can identify more predictive disease models, potentially preventing clinical trials starting based on unreliable data. Ethics committees have a unique opportunity to incentivise higher standards since an unfavourable assessment can prevent poorly designed experiments from even starting. The request of a more detailed translational rationale for each model choice (e.g. by requiring models are evaluated with FIMD), as well as the enforcement of reporting guidelines, can act as gatekeepers for flawed study designs and improve the risk-benefit analysis significantly.

Funders can require the use of systematic reviews and meta-analyses and a thorough assessment of the translational relevance of selected animal models (e.g. FIMD). Journal editors and reviewers must actively enforce reporting guidelines for ongoing submissions as endorsing them does not improve compliance. Regulatory agencies can shape the drug development landscape significantly by, for instance, updating the IB guidelines by requiring a more extensive translational rationale for each animal model employed. Scientific advice can be used as a platform to discuss translational considerations early in development.

Larger companies can perform a thorough assessment of preclinical data of internal and external assets using FIMD, systematic review and meta-analysis, and the IB-derisk. At the same time, small and medium enterprises can provide data in these formats to support their development plan. Ultimately, the selection of more predictive disease models will lead to more successful clinical trials, increasing the benefit and reducing the risks to patients, and lower development costs.

A positive side-effect of these strategies is the increased scrutiny of design and animal model choices. Instead of a status-quo based on tradition and replication of poor practices, we can move forward to an inquisitive and evidence-based modus operandi. This change of culture is sorely needed in both academic and industrial

institutions. A shift toward a stricter approach – more similar to clinical trials, from beginning to end, is warranted. The harmonisation of requirements across stakeholders will be crucial for a successful change of mindset.

Fortunately, some of these changes are already underway. In the past decade, scientific organisations, such as the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3R), have continuously subsidised research aimed at improving the internal validity of animal studies. The Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES) and the Systematic Review Center for Laboratory animal Experimentation (SYRCLE) are relevant examples of groups supported by the NC3Rs that promote systematic reviews and meta-analyses as tools to diagnose the problems with current animal research – and propose solutions based on these results. Moreover, the Food and Drug Administration (FDA) may also expand their current framework called 'the Animal Rule', which currently covers the approval of products treating diseases for which human efficacy testing is neither feasible nor ethical, to qualify all disease models. Together with tools that assess disease models' external validity, like those presented in this thesis, these initiatives may lead to a significant improvement in pharmaceutical R&D productivity.

In the foreseeable future, animal research is unlikely to be eliminated. The criticism of the present state of affairs of preclinical research is indeed justifiable. Our efforts must be focused on improving the robustness of animal data generated now. We already have tools available to address most, if not all, internal and external validity concerns. Only a thorough assessment of higher-quality animal data will determine whether animal research is still a valid paradigm in drug development. By promoting and adopting high-quality study design and reporting as well as a thorough assessment of the translatability of drug efficacy of animal models of disease, we will have robust data to improve the current animal research paradigm.

Samenvatting

Anders dan in de jaren negentig, toen 40% van de geneesmiddelen bij de mens faalde door problemen met farmacokinetiek en biobeschikbaarheid, zijn nu de belangrijkste redenen voor het falen de klinische werkzaamheid, veiligheid en commerciele motivaties, in deze volgorde. Nieuwe geneesmiddelen vinden op basis van dierstudieshun weg naar de kliniek. Dit suggereert dat bij onvoldoende werkzaamheid of veiligheid in de mens dat diermodellen niet voorspellend genoeg zijn. Slechte reproduceerbaarheid wordt vaak genoemd als de primaire oorzaak voor dit hoge percentage van falen. Daarom zijn de meeste inspanningen tot nu toe gericht op het verbeteren van kwaliteit in de opzet, uitvoering en rapportage van dierproeven (interne validiteit). Momenteel is een verbetering van de interne validiteit afhankelijk van de implementatie in plaats van de ontwikkeling van richtlijnen en aanbevelingen.

De evaluatie van diermodellen zelf (externe validiteit) wordt vaak over het hoofd gezien. De huidige manieren om de externe validiteit van ziektemodellen te beoordelen hebben verschillende beperkingen. Ze zijn niet gestandaardiseerd, kunnen door elke gebruiker verschillend worden geïnterpreteerd, sluiten een eenvoudige vergelijking tussen diermodellen uit of missen een diepgaandere analyse van pathofysiologische kenmerken die relevant zijn voor de werkzaamheid van het geneesmiddel. Desalniettemin is het alleen door het vaststellen van de essentiële informatie waarmee diermodellen kunnen worden geëvalueerd dat we modellen kunnen selecteren die de menselijke respons beter kunnen voorspellen. Daarom is er behoefte aan nieuwe manieren om onderscheid te maken tussen potentieel voorspellende diermodellen van ziekten.

In dit proefschrift hebben we nieuwe manieren ontwikkeld om onderscheid te maken tussen potentieel voorspellende diermodellen. We hebben dstandaarden gedefinieerd die als leidraad dienen voor de beoordeling, vergelijking en validatie van diermodellen die gebruikt worden om de voorlopige werkzaamheid van nieuwe geneesmiddelen te testen. We baseerden onze aanpak op het schetsen van de relaties tussen dierlijke en menselijke ziekten op fysiologisch, farmacologisch en therapeutisch niveau.

In **hoofdstuk 2** presenteren we de ontwikkeling van het Framework to Identify Models of Disease (FIMD). We hebben in de literatuur acht domeinen geïdentificeerd die door onderzoekers worden gebruikt bij de validatie en optimalisatie van diermodellen van ziekten: Epidemiologie, Symptomatologie en Natuurlijke historie (SNH), Genetica, Biochemie, Etiologie, Histologie, Farmacologie en Eindpunten. Op basis van deze domeinen werden vragen opgesteld om de overeenkomsten

tussen dierlijke en menselijke toestand? te beoordelen. We hebben een wegings- en scoresysteem opgenomen dat alle domeinen gelijk gewogen heeft. Bovendien bevat FIMD metingen van de vergelijkbaarheid tussen diermodellen, onzekerheid over de validatie, de kwaliteit van de rapportering en het risico op vertekening. FIMD is aangevuld met instructies om de toepassing ervan te vergemakkelijken bij de beoordeling, validatie en vergelijking van diermodellen van ziekten.

In **hoofdstuk 3** gaan we dieper in op de validatie van FIMD, die een vereenvoudigde versie van het framework (pilotstudie) voor twee diermodellen van type 2 diabetes (T2D) en een volledige validatie voor twee modellen van Duchenne spierdystrofie (DMD) omvatte. In de pilotstudie suggereerden onze resultaten van de T2D-modellen Zucker Diabetic Fatty (ZDF) rat en db/db muis dat er slechts kleine verschillen zijn tussen deze modellen, hoewel het farmacologische domein niet genoeg studies had. De volledige validatie van de DMD-modellen mdx muis en Golden Retriever Muscular Dystrophy (GRMD) hond daarentegen toonde significante verschillen in vijf domeinen: Epidemiologie, SNH, Histologie, Farmacologie en Eindpunten. Terwijl de GRMD hond beter scoorde in de eerste drie, was de mdx muis superieur in de laatste twee. Toch kan de hogere score van de mdx muis gedeeltelijk worden toegeschreven aan de afwezigheid van studies bij de hond, wat leidt tot een lagere score in de domeinen Farmacologie en Eindpunten. De kwaliteit van de rapportage was slecht en het risico op vertekening kon in de meeste gevallen niet worden beoordeeld. Met de volledige validatie van de DMD-modellen als voorbeeld, introduceren we een stapsgewijze aanpak om een volledige validatie van elk diermodel van de ziekte uit te voeren.

In **hoofdstuk 4** hebben we de resultaten van de pilotstudie van FIMD in T2D, die slechts kleine verschillen tussen de ZDF-rat en de db/db-muis lieten zien, verder onderzocht. Aangezien FIMD de overeenkomsten tussen dier- en humanegegevens op therapeutisch niveau niet volledig heeft onderzocht, hebben we een systematische review en meta-analyse van dierstudies uitgevoerd om de resultaten van de pilotstudie te valideren. Het glucoseverlagende effect van goedgekeurde geneesmiddelen op de geglyceerde hemoglobine A1c (HbA1c) werd in beide modellen vergeleken. Onze bevindingen gaven aan dat de kleine verschillen tussen de ZDF-rat en de db/db-muis niet resulteerden in verschillen in het therapeutische effect van de geneesmiddelen, met uitzondering van exenatide. Vergelijkbaar met de resultaten van de pilotstudie kon het risico op vertekening niet worden geëvalueerd voor de meeste studies, vanwege de lage kwaliteit van de studierapportages. Deze systematische review en meta-analyse legt de basis voor de toepassing van deze methodologie om gegevens over de werkzaamheid van geneesmiddelen in dieren en mensen te vergelijken.

In **hoofdstuk 5** onderzoeken we preklinische farmacokinetische en farmacodynamische voorspellingen in de vroege klinische ontwikkeling om ziektemodellen op farmacologisch niveau te beoordelen. De Investigator's Brochure (IB) is een belangrijk document die relevante non-klinische informatie samenvat vóór de first-in-human studies. Deze is gebruikt als bronmateriaal. We hebben de IB-derisk tool (ontwikkeld bij het Centrum voor Humane Geneesmiddelenonderzoek - CHDR, Leiden, NL) gebruikt om dierlijke en menselijke farmacokinetische en farmacodynamische gegevens te integreren. Onze resultaten suggereren dat, hoewel diermodellen mogelijkerwijs een voorspelling kunnen doen over de goed verdraagbare en farmacologisch actieve bereik inde mens, er nog veel ruimte is voor verbetering. De rapportagekwaliteit van het IB was vergelijkbaar met onze eerdere beoordelingen in hoofdstuk 3 en 4, wat aangeeft dat slechte rapportage zich niet beperkt tot de wetenschappelijke literatuur. We stellen strengere eisen voor m.b.t. het ontwerp, de rapportage en de uitvoering van dierproeven, een bredere verkenning van de blootstelling in effectiviteitsstudies, naast andere maatregelen om het potentieel van het IB om de vertaling van dier naar mens te verbeteren.

Tenslotte wordt in **hoofdstuk 6**, de discussie, elke methode die in dit proefschrift wordt besproken recapituleerd: FIMD, systematische reviews en meta-analyse en de IB-derisk tool. We bespreken hun toepassingen, uitdagingen voor de implementatie en de impact op proefdier onderzoek. Het gebruik van deze nieuwe methoden om onderscheid te maken tussen ziektemodellen kan alleen maar verder gaan in het aanpakken van de hoge uitvalpercentage in de ontwikkeling van geneesmiddelen. Alleen met een gezamenlijke inspanning van onderzoekers in de academische wereld en de industrie, ethische commissies, financiers, regelgevende instanties en de farmaceutische industrie kan de kwaliteit van dierproeven worden verbeterd.

Door toepassing van FIMD, systematische reviews en meta-analyse, en de IB-derisk tool, kunnen onderzoekers beter voorspellende ziektemodellen identificeren, waardoor mogelijk wordt voorkomen dat klinische studies starten op basis van onbetrouwbare gegevens. Ethische commissies hebben hierin een unieke kans om hogere standaarden te stimuleren, omdat een ongunstige beoordeling kan voorkomen dat slecht ontworpen experimenten worden gestart. Het verzoek om een meer gedetailleerde translationele motivatie voor elke modelkeuze (bijvoorbeeld door te eisen dat modellen met FIMD worden geëvalueerd), alsmede de handhaving van rapportagerichtlijnen, kunnen als poortwachter fungeren voor gebrekkige onderzoeksontwerpen en de risico-batenanalyse aanzienlijk verbeteren.

Bovendien, financiers kunnen het gebruik van systematische reviews en meta-analyses en een grondige beoordeling van de translationele relevantie van geselecteerde diermodellen (bv. FIMD) eisen. Tijdschriftredacteuren en beoordelaars moeten

actief toezien op de naleving van de rapportagerichtlijnen van lopende aanvragen, aangezien stimulering daarvan de naleving niet verbetert. Regelgevende instanties kunnen het geneesmiddelenontwikkelingslandschap vormgeven door bijvoorbeeld de IB richtlijnen te actualiseren door een uitgebreidere translationele onderbouwing van elk gebruikt diermodel te eisen. Wetenschappelijk advies kan worden gebruikt als een platform voor het bespreken van de translationele overwegingen in een vroeg stadium van de ontwikkeling.

Grotere bedrijven kunnen een grondige beoordeling van preklinische gegevens van interne en externe middelen uitvoeren met behulp van FIMD, systematische reviews en meta-analyse, en de IB-derisk tool. Tegelijkertijd kunnen kleine en middelgrote ondernemingen gegevens in deze formaten aanleveren ter ondersteuning van hun ontwikkelingsplan. Uiteindelijk zal de selectie van beter voorspellende ziektemodellen leiden tot meer succesvolle klinische studies, waardoor het voordeel en de risico's voor de patiënten toenemen en de ontwikkelingskosten dalen.

Een positief neveneffect van deze strategieën is de toegenomen aandacht voor het ontwerp en de keuze van diermodellen. In plaats van een status-quo gebaseerd op traditie en replicatie van slechte praktijken, kunnen we overgaan tot een onderzoekende en op bewijs gebaseerde modus operandi. Deze cultuuromslag is hard nodig in zowel academische als industriële instellingen. Een verschuiving naar een striktere aanpak - die meer lijkt op klinisch onderzoek, en van begin tot eind is gerechtvaardigd. De harmonisatie van de eisen voor alle belanghebbenden zal van cruciaal belang zijn voor een succesvolle mentaliteitsverandering.

Gelukkig zijn sommige van deze veranderingen al in gang gezet. In het afgelopen decennium hebben wetenschappelijke organisaties, zoals het Nationaal Centrum voor Vervanging, Verfijning en Vermindering van Dieren in Onderzoek (NC3Rs) in het Verenigd Koninkrijk, onderzoek gesubsidieerd dat gericht is op het verbeteren van de interne validiteit van dierproeven. De Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES) in de Universiteit van Edinburgh is een relevante voorbeeld van groepen die door de NC3Rs worden ondersteund. In Nederland, het Systematic Review Center for Laboratory Animal Experimentation (SYRCLE) in de Radboud Universiteit Nijmegen is een soortgelijk initiatief. Beide CAMARADES and SYRCLE bevorderen systematische reviews en meta-analyses als instrumenten om de problemen met het huidige dieronderzoek te diagnosticeren – en op basis van deze resultaten oplossingen voorstellen. Bovendien kan de Food and Drug Administration (FDA) hun huidige raamwerk, "the Animal Rule" genaamd, dat momenteel de goedkeuring van producten voor de behandeling van ziekten waarvoor het testen van de werkzaamheid bij de mens noch haalbaar noch ethisch verantwoord is, uitbreiden om alle ziekten te

kwalificeren die in het kader van het onderzoek bij dieren worden aangetroffen. Samen met instrumenten die de externe validiteit van ziektemodellen beoordelen, zoals die welke in dit proefschrft worden gepresenteerd, kunnen deze initiatieven leiden tot een aanzienlijke verbetering van de productiviteit van farmaceutisch onderzoek en ontwikkeling.

Binnen afzienbare tijd zal proefdieronderzoek waarschijnlijk niet worden geëlimineerd. Maar de kritiek op de huidige stand van zaken van het preklinische onderzoek isgerechtvaardigd. Onze inspanningen moeten gericht zijn op het verbeteren van de robuustheid van de diergegevens die nu worden gegenereerd. We beschikken al over instrumenten om de meeste, zo niet alle, interne en externe validiteitsproblemen aan te pakken. Alleen een grondige beoordeling van hoogkwalitatieve diergegevens zal bepalen of proefdieronderzoek/onderzoek in dieren nog steeds een geldig paradigma is in de ontwikkeling van geneesmiddelen. Door het bevorderen en aannemen van zowel hoogkwalitatieve onderzoeksopzet en -rapportage en een grondige beoordeling van de vertaalbaarheid van medicijneffectiviteit in ziektemodellen, zullen we over robuuste gegevens beschikken om het huidige proefonderzoeksparadigma te verbeteren.

Resumo

Diferentemente dos anos 90, quando 40% dos medicamentos falhavam em humanos devido à questões de farmacocinética e biodisponibilidade, agora as principais razões para o fracasso são a eficácia e segurança clínicas, e comercial, nesta ordem. Uma vez que novos medicamentos frequentemente avançam para a fase clínica com base em estudos em animais, eficácia ou segurança insuficiente em humanos sugere que modelos animais não são suficientemente preditivos. A baixa reprodutibilidade tem sido frequentemente citada como a principal causa desta alta taxa de fracasso. Por esta razão, a maioria dos esforços até agora se concentraram em lidar com a baixa qualidade no planejamento, condução e relato de estudos em animais (validade interna). Atualmente, uma melhoria da validade interna depende da implementação, e não do desenvolvimento, de diretrizes e recomendações.

Por outro lado, a avaliação dos modelos animais em si (validade externa) tem sido muitas vezes negligenciada. As formas atuais de avaliar a validade externa dos modelos animais de doenças têm várias limitações. Elas não são padronizadas, podem ser interpretadas diferentemente por cada usuário, impedem uma comparação fácil entre modelos animais ou carecem de uma análise mais profunda das características fisiopatológicas relevantes para a eficácia das drogas. No entanto, só é possível selecionar modelos animais mais propensos a prever a resposta humana através da definição de quais são as informações essenciais com as quais os modelos animais podem ser avaliados. Portanto, são necessárias novas formas de diferenciar modelos animais de doença potencialmente preditivos.

Nesta tese, desenvolvemos novos modos de diferenciar modelos animais potencialmente preditivos. Definimos os padrões para orientar a avaliação, comparação e validação dos modelos animais usados para testar a eficácia preliminar de novos medicamentos. Baseamos nossa abordagem na delineação das relações entre as doenças animais e humanas a nível fisiológico, farmacológico e terapêutico.

No **Capítulo 2**, apresentamos o desenvolvimento do Framework para Identificar Modelos de Doenças (FIMD). Identificamos na literatura oito domínios utilizados pelos pesquisadores na validação e otimização de modelos animais de doenças: Epidemiologia, Sintomatologia e História Natural (SNH), Genética, Bioquímica, Etiologia, Histologia, Farmacologia e Desfechos. Com base nestes domínios, elaboramos perguntas para avaliar as semelhanças entre a condição animal e humana. Incluímos um sistema de ponderação e pontuação em que todos os domínios recebiam o mesmo peso. Além disso, o FIMD apresenta medidas

de comparabilidade entre modelos animais, incerteza em torno da validação, qualidade dos relatórios e risco de viés. O FIMD é completado com instruções para facilitar sua aplicação na avaliação, validação e comparação de modelos animais de doenças.

No **Capítulo 3**, detalhamos a validação do FIMD, que incluiu uma versão simplificada do framework (estudo piloto) para dois modelos animais de Diabetes Tipo 2 (DT2) e uma validação completa para dois modelos de Distrofia Muscular de Duchenne (DMD). No estudo piloto, os nossos resultados nos modelos de DT2, o rato Zucker Diabetic Fatty (ZDF) e o camundongo db/db, sugerem que existem apenas pequenas diferenças entre eles, embora o domínio da Farmacologia careça de estudos. Por outro lado, a validação completa dos modelos de DMD, o camundongo mdx e o cão Golden Retriever Muscular Dystrophy (GRMD), mostrou diferenças significativas em cinco domínios: Epidemiologia, SNH, Histologia, Farmacologia e Desfechos. Enquanto o cão GRMD teve melhor pontuação nos três primeiros, o camundongo mdx foi superior nos dois últimos. Entretanto, a pontuação mais alta do camundongo mdx pode ser parcialmente atribuída aos poucos estudos no cão, levando a uma pontuação mais baixa nos domínios de Farmacologia e Desfechos. A qualidade do relatório foi baixa, e o risco de viés não pode ser avaliado na maioria dos casos. Usando a validação completa dos modelos de DMD como exemplo, introduzimos uma abordagem por etapas para realizar uma validação completa de qualquer modelo animal de doença.

No **Capítulo 4**, exploramos mais profundamente os resultados do estudo piloto do FIMD em DT2, que mostrou apenas pequenas diferenças entre o rato ZDF e o comundongo db/db. Como o FIMD não investigou completamente a concordância entre dados de animais e humanos a nível terapêutico, realizamos uma revisão sistemática e uma meta-análise dos estudos com animais para validar os resultados do estudo piloto. O efeito de diminuição da glicose de medicamentos aprovados na hemoglobina glicada A1c (HbA1c) foi comparado em ambos os modelos. Nossas descobertas indicam que as poucas distinções entre o rato ZDF e o camundongo db/db não resultam em diferenças no efeito terapêutico de todos os medicamentos, exceto o exenatide. Como nos resultados do estudo piloto, o risco de viés não pode ser avaliado na maioria dos estudos devido à baixa qualidade de descrição dos experimentos. Esta revisão sistemática e meta-análise estabelece a base para a aplicação desta metodologia na comparação de dados de eficácia animal e humana.

No **Capítulo 5**, exploramos as previsões farmacocinéticas e farmacodinâmicas pré-clínicas no desenvolvimento clínico inicial para avaliar modelos de doença no nível farmacológico. A Brochura do Investigador (IBs), um documento fundamental que

resume todas as informações relevantes antes dos primeiros estudos em humanos, foram usados como material de partida. Utilizamos a ferramenta IB-derisk (desenvolvida no Centre for Human Drug Research - CHDR, Leiden, NL) para integrar dados farmacocinéticos e farmacodinâmicos de animais e humanos. Nossos resultados indicam que, embora os modelos animais possam potencialmente prever as faixas humanas bem toleradas e farmacologicamente ativas, há muito espaço para melhorias. A qualidade dos relatórios do IB foi semelhante às nossas avaliações anteriores nos Capítulos 3 e 4, sugerindo que a baixa qualidade da evidência não se restringe à literatura científica. Propomos requisitos mais rigorosos para o planejamento, descrição e condução de experimentos com animais, exploração mais ampla da exposição em estudos de eficácia, entre outras medidas para melhorar o potencial da IB para auxiliar a tradução de animais para humanos.

Finalmente, no **Capítulo 6**, a discussão, recapitulamos cada método debatido nesta tese: FIMD, revisões sistemáticas e meta-análises e a ferramenta IB-derisk. Discutimos suas aplicações, desafios para implementação e impacto na pesquisa com animais. O uso destes novos métodos para diferenciar modelos animais de doenças sozinho tem apenas efeito limitado sobre as altas taxas de atrito no desenvolvimento de medicamentos. Somente com um esforço conjunto, envolvendo pesquisadores da academia e da indústria, comitês de ética, financiadores, agências reguladoras e a indústria farmacêutica, a qualidade da pesquisa animal pode ser melhorada.

Ao aplicar o FIMD, revisões sistemáticas e meta-análises e o IB-derisk, os pesquisadores podem identificar modelos de doenças mais preditivos, possivelmente prevenindo o início de ensaios clínicos com base em dados não confiáveis. Os comitês de ética têm uma oportunidade única para incentivar padrões mais altos, já que uma avaliação desfavorável pode impedir que experimentos mal concebidos sejam sequer começados. A solicitação de uma fundamentação translacional mais detalhada para cada escolha de modelo (por exemplo, exigindo que os modelos sejam avaliados com o FIMD), bem como a aplicação das diretrizes de descrição de experimentos, podem atuar como a triagem de desenhos de estudo com falhas e melhorar significativamente a análise de risco-benefício.

Os financiadores podem exigir o uso de revisões sistemáticas e meta-análises e uma avaliação completa da relevância translacional de modelos animais selecionados (por exemplo, FIMD). Os editores e revisores de periódicos devem aplicar ativamente as diretrizes para a apresentação de relatórios em andamento, pois somente endossá-las não melhora a aderência a estas diretrizes. As agências reguladoras podem moldar significativamente o cenário de desenvolvimento de medicamentos, por exemplo, atualizando as diretrizes da IB, exigindo uma fundamentação translacional mais extensa para cada modelo animal empregado.

O aconselhamento científico pode ser usado como uma plataforma para discutir considerações translacionais no início do desenvolvimento.

As grandes empresas podem realizar uma avaliação completa dos dados pré-clínicos dos ativos internos e externos usando FIMD, revisão sistemáticas e meta-análises, e o IB-derisk. Ao mesmo tempo, as pequenas e médias empresas podem fornecer dados nestes formatos para apoiar seu plano de desenvolvimento. Em última análise, a seleção de modelos de doenças mais preditivos levará a ensaios clínicos mais bem sucedidos, aumentando o benefício e reduzindo os riscos para os pacientes, além de diminuir os custos de desenvolvimento.

Um efeito colateral positivo dessas estratégias é o maior escrutínio das escolhas de desenho de estudo e modelos animais. Em vez de um status quo baseado na tradição e replicação de práticas questionáveis, podemos avançar para um modus operandi inquisitivo e baseado em evidências. Esta mudança de cultura é fundamental, tanto em instituições acadêmicas quanto industriais. Uma mudança em direção a uma abordagem mais estrita – mais semelhante aos ensaios clínicos, do início ao fim – é necessária. A harmonização das exigências entre as partes interessadas será crucial para uma mudança bem sucedida de mentalidade.

Felizmente, algumas dessas mudanças já estão em andamento. Na última década, organizações científicas, como o Centro Nacional de Substituição, Refinamento e Redução de Animais em Pesquisa (NC3R), têm subsidiado continuamente pesquisas que visam melhorar a validade interna dos estudos com animais. A Abordagem Colaborativa para Meta-Análise e Revisão de Dados de Animais de Estudos Experimentais (CAMARADES) e o Centro de Revisão Sistemática para Experimentação Animal em Laboratório (SYRCLE) são exemplos relevantes de grupos apoiados pelos NC3Rs que promovem revisões sistemáticas e meta-análises como ferramentas para diagnosticar os problemas com a pesquisa animal atual – e propor soluções baseadas nestes resultados. Além disso, o Food and Drug Administration (FDA) também pode expandir seu framework, chamado "Animal Rule", que atualmente cobre a aprovação de produtos que tratam doenças para as quais os testes de eficácia humana não são nem viáveis nem éticos, para qualificar todos os modelos de doenças. Junto com ferramentas que avaliam a validade externa dos modelos de doenças, como as apresentadas nesta tese, estas iniciativas podem levar a uma melhoria significativa na produtividade da P&D farmacêuticos.

Num futuro próximo, é pouco provável que a pesquisa com animais seja eliminada. A crítica à situação atual da pesquisa pré-clínica é de fato justificável. Nossos esforços devem ser concentrados em melhorar a robustez dos dados animais gerados atualmente. Já temos ferramentas disponíveis para tratar da maioria, senão de todas, as preocupações de validade interna e externa. Somente uma avaliação

minuciosa dos dados de animais de maior qualidade determinará se a pesquisa com animais ainda é um paradigma válido no desenvolvimento de medicamentos. Ao promover e adotar um desenho e descrição de estudo de alta qualidade, bem como uma avaliação completa da translatabilidade da eficácia de medicamentos em modelos animais de doenças, produziremos dados robustos para melhorar o paradigma atual da pesquisa com animais.

Acknowledgements

Science is done by many hands. The work presented in this thesis was at least partially conceived, performed, written and reviewed by hands and minds other than mine. Some of them hold a PhD degree. Others have never set foot in an academic institution. My acknowledgements start with the most important of them all.

Mamãe, essa tese não existiria sem você. A trajetória que percorri até esse momento foi longa e dura, mas só foi possível porque você me proporcionou as condições para tal. Obrigado por ter repetido exaustivamente que o estudo era o único caminho. Obrigado pela preocupação constante, contatos diários e amor inesgotável. Obrigado por ter superado a morte do seu grande amor, meu pai, para continuar a nos encorajar a seguir adiante. Obrigado por sempre acreditar e reforçar que eu alcançaria os meus sonhos – mesmo que eles significassem um oceano entre nós. Você me deu os recursos que me possibilitaram chegar aqui. Você pode nunca ter concluído seus estudos, mas seu papel na conclusão dos meus é indiscutível. Mais uma vez, esse diploma também é seu, Drª. Valnice!

Euler, obrigado por sempre estar disponível para resolver os meus problemas técnicos e práticos, do Brasil e da Holanda. Obrigado por acalmar e cuidar da pessoa mais importante no mundo para gente. Às minhas tias, tios, primas e primos, obrigado por serem meus beijos e abraços para essas duas pessoas que são a minha fundação. Obrigado também por sempre acreditarem que esse dia chegaria.

Huub, thank you for providing your expertise to the project. Ellen, thank you for being present, responsive and committed to this research. Thank you for solving the practical matters and listening regardless of your opinion. Peter, thank you for always providing the second perspective throughout the development of this thesis. In between your ever-valuable insights, I genuinely appreciated our conversations about the fantasy worlds far away from the reality of animal experimentation. Thank you for following up on the idea that led to my position. Wouter, thank you for always timely reviewing my output, without ever losing kindness out of sight. To all of you, thank you for allowing me to come back to the country where I truly feel at home. You made a dream come true.

Carlijn, thank you for sticking with me through our short collaboration. Your responsiveness and sharp comments were fundamental to successfully start and finalise the systematic review and meta-analysis. Joop, thank you for being so responsive and open to collaborating. Your input was precious for the design of the project and interpretation of the results. Rik, thank you for taking me in and for always making time to discuss all the matters at hand. Although each of you participated in a single project, you surely left your mark in this thesis, and in me.

Wim, thank you for bringing clarity and perspective when I most needed. Barbara, thank you for the lovely talks and lekkere knuffelen all these years. Without your support in resolving the countless administrative and practical matters, this thesis would be far from complete. Antoinette, thank you for advising me in the most delicate issues. Saskia, thank you for listening to me and pointing me to the people that could help me when I most needed it. Toine, thank you for lending me your ear and taking the time to help me realise this thesis, even if indirectly.

Désirée, I could write a whole chapter just for you. Since the beginning, you have been a kind, attentive, caring and competent colleague. Soon enough, you became a friend. Thank you for the hardwork and unyielding support. You are the reason why I stayed sane in the most stressful moment of my life. I owe this thesis to you. I feel so privileged to have you around I could never express in a paragraph. Thank you for being my rock at DDW and beyond.

Edwin, thank you for all these years of mentoring. Thank you for showing me the path to become who I want to become. Boris, thank you for your time, attention and knowledge. Thank you for the advice on how to deal with every challenge I have ever encountered. Thank you for helping me see the trees, but also the forest for the trees.

To our Lunch Group (Lies, Sjaak, Erik, Carl, Charis, Jeffrey), thank you for the companionship, the trust and ease of laugh in our first years. Karina, thank you for being there at every step of the way. Thank you for being responsive, kind and also for trying to help me with the German taxes. The PhD brought me a few gifts, but you came with the worst package, Maria. On Day 1, I could not fathom you would become such an essential person in my life. Thank you for the drinks, laughs, shade, kind and harsh words. Thank you for bringing another point of view on every topic we ever discussed. Thank you for being there at so many crucial moments – good and bad.

Fatema, thank you for trusting me and giving me the time and space I needed to fully dedicate myself to work at the Access to Medicine Foundation (ATMF). You, and everyone at ATMF (especially Dulce, Mireille, Moska, Ruth and Teresa), allowed me to experience a healthy working environment. Thank you for reminding me of the beauty of working with dedicated, competent and kind people.

Thank you to all my friends, who stood unwavering even when I faltered. Thank you, Adam, Adrian, Adrie, Álvaro, Ana, Antonio, Cesc, Daniel, Dirk, Eder, Felipe, Fernanda, Filipe, Gabi, Imad, Jéssica, Juanmi, Luiz, Mikel, Paty, Raquel, Rico, Rubén, Salvatore, Sebastian, Unai and Zaki. Filipi, thank you for the advice and help with designer aspects of the thesis. Filipi, thank you for helping me with design tips and an always positive outlook. Cindy and Adriano, thank for your supporting even before this

position was offered to me. Thank you for spending your holiday time to rehearse the pitch that led to this PhD. Karim, thank you for existing in my life, for showing me a kind and chaotic way to see and do things. Thank you for giving me the privilege of experiencing your genius in action. Dimitri, thank you for comforting me in one of my toughest nights, and for never ceasing to amaze me with your wit, compassion and smile. Eti, my officially declared role model, you have encouraged me every single time I thought I would not make it. Thank you for always being the reassuring voice saying I could do it, regardless of how dire my situation seemed. Every day with you is a reminder of the person I strive to become. Thank you.

Delfina, obrigado por me ouvir e sempre oferecer um caminho iluminado para seguir. Você torna a minha vida mais leve e feliz.

Elaine, you are my anti-inner saboteur. For the past seven years, you have been my most generous supporter. An ocean of distance has never come in your way to show me your love, your care and your confidence that I would achieve my dreams. Our connection lacks an objective explanation, but I have come with terms that we do not need one. You are a constant reminder that our starting point only tells the beginning of the story. We get to choose who we want to become, where we want to go and how we want to move forward. Thank you for so often acting also as a supervisor, reviewer and editor. Thanks for sharing your life with me through breias, companionship, laughter and love.

Robbie, you came around in the middle of a storm and yet, you persevered. You reminded me, every step of the way, this thesis would be finished. You understood the lack of time and provided the sorely needed hugs and comfort in the most challenging times. Thank you for consistently making me feel protected and cared for. Thank you for showing me the road to a healthy and thriving relationship may be rocky, but it is certainly worth the drive if you have the 'right person'. You are my 'right person'. Thank you for showing me real love. This is the first of the many events we will celebrate together, liefde!

And to all the others that may not have been cited by name, but unquestionably supported this thesis: thank you. I am forever indebted to you.

Curriculum Vitae

Guilherme S. Ferreira ("Gui") was born in Rio de Janeiro in 1991. He started in science in 2006, when he joined the Polytechnic School of Health Joaquim Venâncio (ESPJV) in Oswaldo Cruz Foundation (Fiocruz) to do his high-school combined with technical education in Health Biodiagnostic Laboratory. His high-school thesis was entitled 'Biotechnology as a commercial reality: the bioeconomy, the access to drugs and the Research and Development crisis". He then started his Pharmacy studies at the Federal Fluminense University (UFF), in which he was involved in research in Cellular Biology, Physical-Chemistry and Community Pharmacy. His first industry internship was in Microbiological Quality Control at Farmoquímica (BR), followed by Patents at an intellectual property legal firm and Analytical Development at Merck (BR). In 2012, he received a scholarship for excellent students from the Brazilian government to study one year in the Netherlands ("Science without Borders"). Here, he studied one year of the Master in Drug Innovation, doing his minor internship in Clinical Development at Kinesis Pharma (NL). After returning to Brazil, he did a one-year internship in Regulatory Affairs at Merck S/A, which led him to move in 2015 to the headquarters in Darmstadt for an internship in Pharmacoeconomics and Outcomes Research. There, he also joined the Clinical Pharmacology department for 5 months before returning to Brazil to finish his studies. After graduating in April 2016, his curiosity about how to facilitate the transition from preclinical to clinical development led to him joining the PhD programme in Drug Innovation at Utrecht University under supervision of Huub Schellekens, Elle Moors, Peter van Meer and Wouter Boon. Throughout most of his PhD, he also worked as an external consultant for Merck (DE) in Pharmacoeconomics and for Curare Consulting in Clinical Pharmacology and Translational Science. The results of his PhD are presented in this thesis. He has worked as a Researcher at the Access to Medicine Foundation (NL) and is now a Clinical Development Consultant at 3D PharmXchange (NL).

List of Publications

This Thesis

Levelling the translational gap for animal to human efficacy data. Ferreira GS, Veening-Griffioen DH, Boon WPC, Hooijmans C, Moors EHM, van Meer PJK. Animals, 2020, DOI:10.3390/ani10071199.

Comparison of drug efficacy in two animal models of type 2 diabetes: a systematic review and meta-analysis. Ferreira GS, Veening-Griffioen DH, Boon WPC, Hooijmans C, Moors EHM, Schellekens H, van Meer PJK. European Journal of Pharmacology, 2020, DOI: 10.1016/j.ejphar.2020.173153.

A standardised framework to identify optimal animal models for efficacy assessment in drug development. Ferreira GS, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, Schellekens H, van Meer PJK. Plos One, 2019, DOI: 10.1371/journal.pone.0220325

Other Publications

Pathways in the Drug Development for Alzheimer's Disease (1906-2016): A Bibliometric Study. Schilder IPA, Veening-Griffioen DH, **Ferreira GS**, van Meer PJK, Gispen-de Wied CC, Schellekens H, Boon WPC, Moors EHM. Journal of Scientometric Research, 2020, DOI: 10.5530/jscires.9.3.x.

Tradition, not science, is the basis of animal model selection in translational and applied research. Veening-Griffioen DH, **Ferreira GS**, Boon WPC, Gispen-de Wied CC, Schellekens H, Moors EHM, van Meer PJK. ALTEX, 2020, DOI: 10.14573/altex.2003301.

Are some animal models more equal than others? A case study on the translational value of animal models of efficacy for Alzheimer's disease. D.H. Veening-Griffioen, **G.S. Ferreira**, P. van Meer, W. Boon, C.C. Wied, E. Moors, H. Schellekens. European Journal of Pharmacology, 2019, DOI: 10.1016/j.ejphar.2019.172524.