

Full length article

## Comparison of drug efficacy in two animal models of type 2 diabetes: A systematic review and meta-analysis



Guilherme S. Ferreira<sup>a,\*</sup>, Désirée H. Veening-Griffioen<sup>a</sup>, Wouter P.C. Boon<sup>b</sup>,  
Carlijn R. Hooijmans<sup>c,d</sup>, Ellen H.M. Moors<sup>b</sup>, Huub Schellekens<sup>a</sup>, Peter J.K. van Meer<sup>a,e</sup>

<sup>a</sup> Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands

<sup>b</sup> Copernicus Institute of Sustainable Development, Innovation Studies, Utrecht University, Utrecht, the Netherlands

<sup>c</sup> Department for Health Evidence Unit SYRCLE, Radboud University Medical Centre, the Netherlands

<sup>d</sup> Department of Anesthesiology, Radboud University Medical Centre, Nijmegen, the Netherlands

<sup>e</sup> Medicines Evaluation Board, Utrecht, the Netherlands

### ARTICLE INFO

#### Keywords:

animal model  
Drug development  
Type 2 diabetes  
Translational research  
Systematic review  
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 July 3, 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.

### 1. Introduction

The high attrition rate in drug development has been a topic of debate for over a decade (Kola and Landis, 2004). One of the reasons suggested for such low odds to reach the market is the poor translation of animal to clinical data (Kola and Landis, 2004; Schulz et al., 2016). Efficacy is the leading cause of failure for more than half of all drugs in development, followed by commercial reasons and safety (Hay et al., 2014; Hwang et al., 2016).

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%) (Hay et al., 2014). 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 (International Diabetes Federation, 2017; OCDE/EU, 2018).

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 (Ferreira et al., 2019a, 2019b). 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  $\beta$ -cell transcription that contributes to the diabetic phenotype (Wang et al., 2014).

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

\* Corresponding author.

E-mail address: [g.ferreira@uu.nl](mailto:g.ferreira@uu.nl) (G.S. Ferreira).

<https://doi.org/10.1016/j.ejphar.2020.173153>

Received 24 January 2020; Received in revised form 8 April 2020; Accepted 22 April 2020

Available online 28 April 2020

0014-2999/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(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 (EMA, 2018; FDA, 2008; Nathan et al., 2008, 2007). Additionally, this systematic review can substantiate the basis for a more quantitative approach to discriminate between animal models in any therapeutic area.

## 2. 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.

### 2.1. Search strategy and paper selection

We searched PubMed and Embase on July 3, 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. 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 July 3, 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 (Hummel et al., 1972). GSF and DVG independently screened articles by title and abstract using Rayyan — a web and mobile app for systematic reviews (Ouzzani et al., 2016). 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.

### 2.2. 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 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 (Rohatgi, 2019). 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.

### 2.3. Assessment of methodological quality and risk of bias

GSF and DVG independently performed the quality assessment using SYRCLE's risk of bias tool (Hooijmans et al., 2014). 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.

### 2.4. 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:

$$S. D. = S. E. M. \times \sqrt{n}$$

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 (Higgins et al., 2003).

### 2.5. 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.

## 2.6. 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.

## 3. 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.

### 3.1. Study selection process

The search on PubMed and Embase retrieved 3427 and 4914 publications, respectively (8341 total). After removal of duplicates, 5405 abstracts were assessed (35.2% duplicates), and 1073 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.

### 3.2. Study characteristics

We provide a table with the individual study characteristics in Supporting Information S2. 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 (The Jackson Laboratories, n.d.; “The Zucker Diabetic Fatty (ZDF) Rat: Diet Evaluation Study for the Induction of Type 2 Diabetes in Obese Female ZDF Rats,” n.d.). 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%).

### 3.3. 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.

### 3.4. Meta-analysis of glucose-lowering effect on HbA1c

Due to missing data, we excluded three papers (Jiang et al., 2013; Peterson, 1994; Tullin et al., 2012) 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):  $-1.74$  [ $-2.07, -1.41$ ],  $n = 29$ ,  $I^2 = 51\%$ ),  $P = 0.001$ . There was no difference between sexes ( $P = 0.09$ ).

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  $P = 0.03$ .

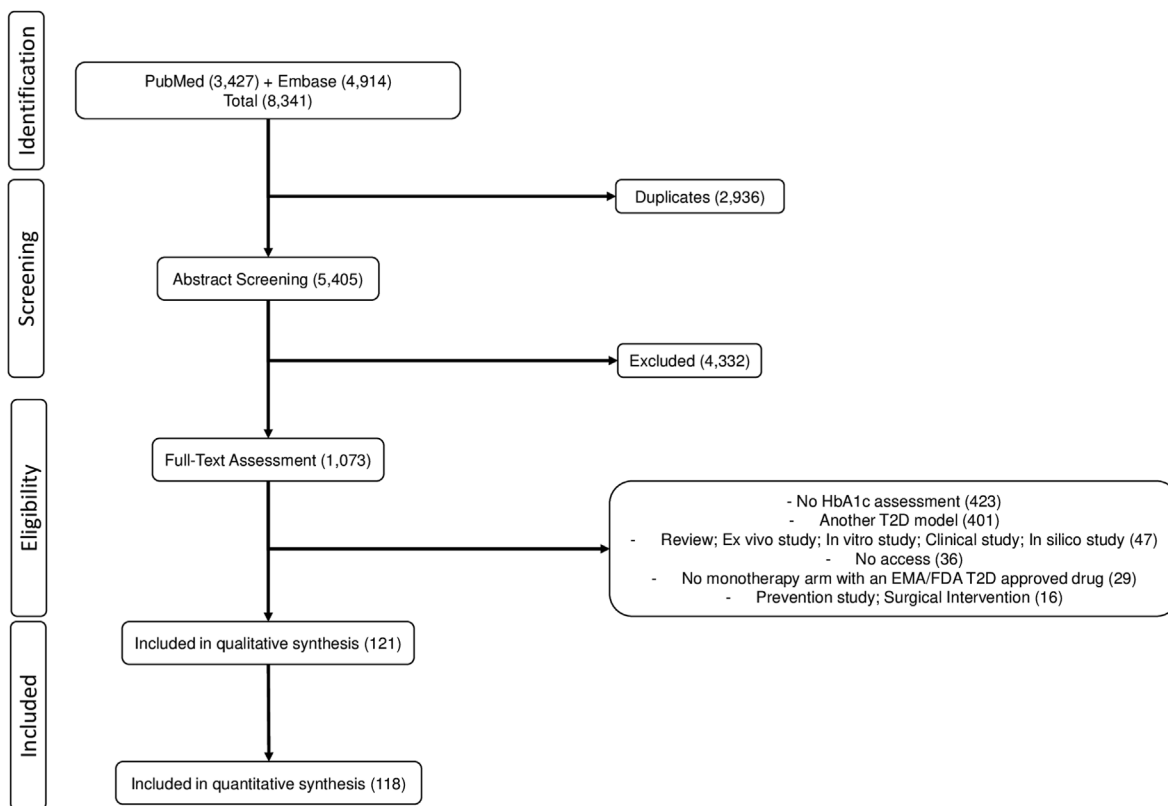


Fig. 1. Flowchart of the study selection process.

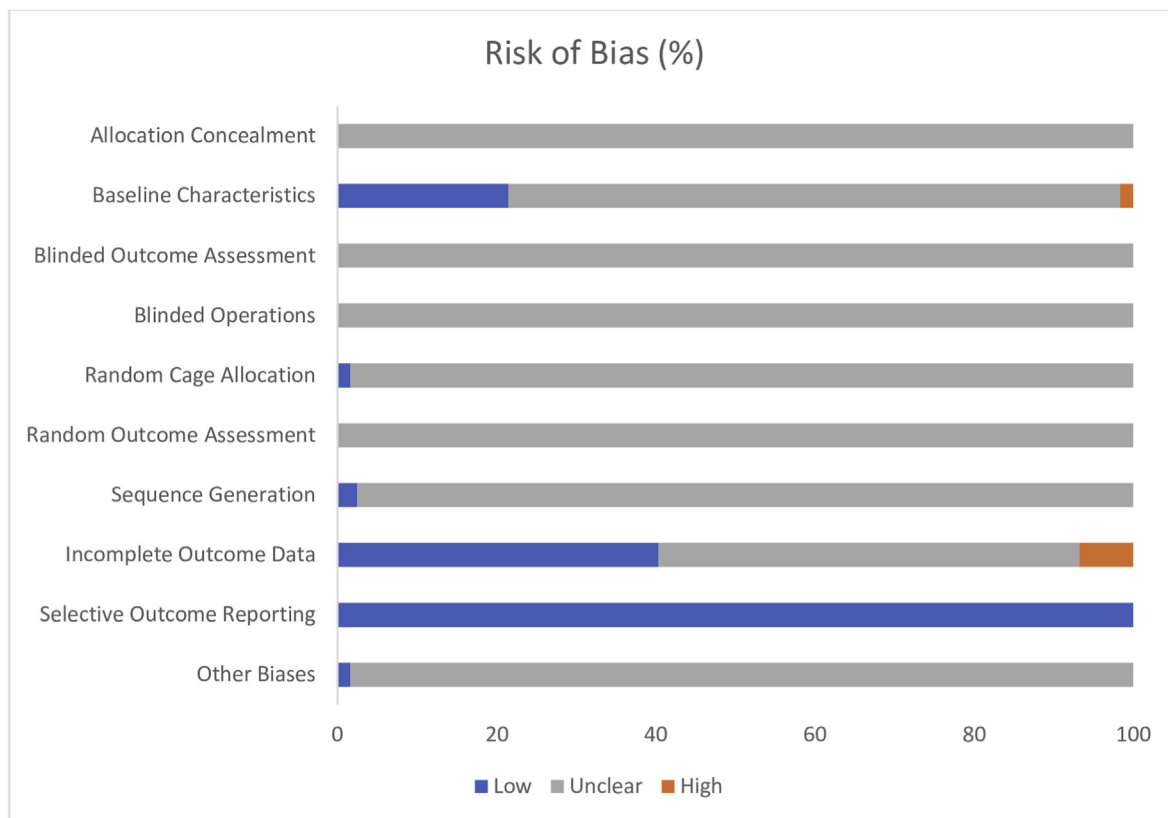


Fig. 2. Results of the risk of bias assessment (n = 121) according to the SYRCLE Risk of Bias Tool.

**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)	
<b>α-Glucosidase Inhibitors</b>					
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
<b>Bile Acid-Binding Resins</b>					
Colesevelam			1.56 [0.92, 2.20]	1 (1)	
<b>DPP-IV Inhibitors</b>					
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)	
<b>GLP-1 Agonists</b>					
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
<b>SGLT-2 Inhibitors</b>					
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)	
<b>Sulphonylureas</b>					
Glibenclamide			-1.03 [-2.81, 0.75]	2 (2)	
Glimepiride	-0.94 [-2.13, 0.25]	2 (2)	-0.23 [-1.09, 0.62]	1 (1)	
<b>Thiazolidinediones</b>					
Thiazolidinediones	-3.63 [-4.29, -2.98]	28 (33)	-2.07 [-3.63, -0.51]	1 (1)	P = 0.74
Pioglitazone	-3.05 [-4.04, -2.06]	8 (8)	-3.98 [-5.38, -2.59]	12 (12)	P = 0.84
Rosiglitazone	-3.88 [-4.70, -3.06]	20 (25)	-3.10 [-4.69, -1.51]	7 (7)	P = 0.37
			-5.19 [-7.92, -2.47]	5 (5)	

CI – confidence interval; DPP-IV - Dipeptidyl Peptidase-4; ES – effect size; GLP-1 – Glucagon-Like Peptide-1; SGLT-2 – Sodium/Glucose Cotransporter-2.

### 3.4.1. 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 ( $P = 0.38$ ).

### 3.4.2. 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).

## 4. 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 (Dhuria et al., 2015). 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. (Ainge et al., 2011; Ranasinghe et al., 2012; Saleh et al., 2018)). 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 (Varga et al., 2015). 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 (Ferreira et al., 2019a). 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

**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.

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 quantitatively 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 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.

#### 4.1. 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 previous literature – animal research is still poorly reported, and often essential details regarding the used methodology are not reported (Bebarta et al., 2003; Begley and Ellis, 2012; Hooijmans et al., 2019). 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 (Wei et al., 2016). 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 (Green, 2015; Sena et al., 2010). However, most drugs and many drug classes had fewer than 10 comparisons and could not be assessed for one or both models.

## 5. 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.

### CRedit authorship contribution statement

**Guilherme S. Ferreira:** Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Désirée H. Veening-Griffioen:** Conceptualization, Investigation, Writing - review & editing. **Wouter P.C. Boon:** Conceptualization, Writing - review & editing, Supervision. **Carlijn R. Hooijmans:** Methodology, Formal analysis, Resources, Writing - review & editing, Supervision. **Ellen H.M. Moors:** Conceptualization, Writing - review & editing, Supervision, Funding acquisition. **Huub Schellekens:** Conceptualization, Supervision, Funding acquisition. **Peter J.K. van Meer:** Conceptualization, Writing - review & editing, Supervision, Project administration, Funding acquisition.

### Acknowledgements

This work was supported by the Dutch Ministries of Agriculture, Nature and Food Quality (MLNV) and Health, Welfare and Sport (MVWS); the Medicines Evaluation Board (CBG); and The Netherlands Organisation for Health Research and Development (ZonMw, grant number 114024136). We want to thank Alice Tillema for her support in optimising the search strategy for PubMed.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejphar.2020.173153>.

## References

- The Zucker Diabetic Fatty (ZDF) Rat: Diet Evaluation Study for the Induction of Type 2 Diabetes in Obese Female ZDF Rats [WWW Document], n.d. URL <https://www.criver.com/sites/default/files/Technical%20Resources/Diet%20Evaluation%20Study%20for%20the%20Induction%20of%20Type%20%20Diabetes%20in%20Obese%20Female%20ZDF%20Rats.pdf> (accessed 9.12.19).
- Ainge, H., Thompson, C., Ozanne, S.E., Rooney, K.B., 2011. A systematic review on animal models of maternal high fat feeding and offspring glycaemic control. *Int. J. Obes.* 35, 325–335. <https://doi.org/10.1038/ijo.2010.149>.
- Bebarta, V., Luyten, D., Heard, K., 2003. Emergency medicine animal research: does use of randomization and blinding affect the results? *Acad. Emerg. Med.* 10, 684–687. <https://doi.org/10.1111/j.1553-2712.2003.tb00056.x>.
- Begley, C.G., Ellis, L.M., 2012. Drug development: raise standards for preclinical cancer research. *Nature* 483, 531–533. <https://doi.org/10.1038/483531a>.
- Dhuria, R.S., Singh, G., Kaur, A., Kaur, R., Kaur, T., 2015. Current status and patent prospective of animal models in diabetic research. *Adv. Biomed. Res.* 4, 117. <https://doi.org/10.4103/2277-9175.157847>.
- EMA, 2018. Guideline on Clinical Investigation of Medicinal Products in 5 the Treatment or Prevention of Diabetes Mellitus.
- Ferreira, G.S., Veening-Griffioen, D.H., Boon, W.P.C., Moors, E.H.M., Gispén-de Wied, C.C., Schellekens, H., van Meer, P.J.K., 2019a. A standardised framework to identify optimal animal models for efficacy assessment in drug development. *PLoS One* 14, e0218014. <https://doi.org/10.1371/journal.pone.0218014>.
- Ferreira, G.S., Veening-Griffioen, D.H., Boon, W.P.C., Moors, E.H.M., Gispén-de Wied, C.C., Schellekens, H., van Meer, P.J.K., 2019b. Correction: a standardised framework to identify optimal animal models for efficacy assessment in drug development. *PLoS One* 14, e0220325. <https://doi.org/10.1371/journal.pone.0220325>.
- FDA, 2008. Guidance for Industry Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention.
- Green, S.B., 2015. 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* 16, 53. <https://doi.org/10.1186/s12910-015-0043-7>.
- Hay, M., Thomas, D.W., Craighead, J.L., Economides, C., Rosenthal, J., 2014. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* 32, 40–51. <https://doi.org/10.1038/nbt.2786>.
- Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. *BMJ* 327, 557–560. <https://doi.org/10.1136/bmj.327.7414.557>.
- Hooijmans, C.R., Rovers, M.M., de Vries, R.B.M., Leenaars, M., Ritskes-Hoitinga, M., Langendam, M.W., 2014. SYRCL's risk of bias tool for animal studies. *BMC Med. Res. Methodol.* 14, 43. <https://doi.org/10.1186/1471-2288-14-43>.
- Hooijmans, C.R., Draper, D., Ergün, M., Scheffer, G.J., 2019. The effect of analgesics on stimulus evoked pain-like behaviour in animal models for chemotherapy induced peripheral neuropathy- a meta-analysis. *Sci. Rep.* 9, 17549. <https://doi.org/10.1038/s41598-019-54152-8>.
- Hummel, K.P., Coleman, D.L., Lane, P.W., 1972. 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.* 7, 1–13. <https://doi.org/10.1007/BF00487005>.
- Hwang, T.J., Carpenter, D., Lauffenburger, J.C., Wang, B., Franklin, J.M., Kesselheim, A.S., 2016. Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Intern Med* 176, 1826–1833. <https://doi.org/10.1001/jamainternmed.2016.6008>.
- International Diabetes Federation, 2017. IDF Diabetes Atlas, 8th. .
- The Jackson Laboratories, n.d. MOUSE STRAIN DATASHEET - 000642 [WWW Document]. MOUSE STRAIN DATASHEET - 000642. URL <https://www.jax.org/strain/000642> (accessed 9.12.19).
- Jiang, B., Guo, S., Shi, D., Guo, C., Wang, T., 2013. 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. *Eur. J. Med. Chem.* 64, 129–136.
- Kola, I., Landis, J., 2004. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–715. <https://doi.org/10.1038/nrd1470>.
- Nathan, D.M., Turgeon, H., Regan, S., 2007. Relationship between glycated haemoglobin levels and mean glucose levels over time. *Diabetologia* 50, 2239–2244. <https://doi.org/10.1007/s00125-007-0803-0>.
- Nathan, D.M., Kuenen, J., Borg, R., Zheng, H., Schoenfeld, D., Heine, R.J., A1c-Derived Average Glucose Study Group, 2008. Translating the A1C assay into estimated average glucose values. *Diabetes Care* 31, 1473–1478. <https://doi.org/10.2337/dc08-0545>.
- OCDE/EU, 2018. Health at a Glance: Europe 2018: State of the Health in the EU Cycle.
- Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan- a web and mobile app for systematic reviews. *Syst. Rev.* 5, 210. <https://doi.org/10.1186/s13643-016-0384-4>.
- Peterson, R.G., 1994. alpha-Glucosidase inhibitors in diabetes: lessons from animal studies. *Eur. J. Clin. Invest.* 24, 11–18.
- Ranasinghe, P., Jayawardana, R., Galappaththy, P., Constantine, G.R., de Vas Gunawardana, N., Katulanda, P., 2012. 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. *Diabet. Med.* 29, 1480–1492. <https://doi.org/10.1111/j.1464-5491.2012.03718.x>.
- Rohatgi, A., 2019. WebPlotDigitizer. accessed 1.8.19. <https://automeris.io/WebPlotDigitizer>.
- Saleh, F., Itani, L., Calugi, S., Grave, R.D., El Ghoch, M., 2018. Adipose-derived mesenchymal stem cells in the treatment of obesity: a systematic review of longitudinal studies on preclinical evidence. *CSCR* 13, 466–475. <https://doi.org/10.2174/1574888X13666180515160008>.
- Schulz, J.B., Cookson, M.R., Hausmann, L., 2016. The impact of fraudulent and irreproducible data to the translational research crisis - solutions and implementation. *J. Neurochem.* 139 (Suppl. 2), 253–270. <https://doi.org/10.1111/jnc.13844>.
- Sena, E.S., van der Worp, H.B., Bath, P.M.W., Howells, D.W., Macleod, M.R., 2010. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol.* 8, e1000344. <https://doi.org/10.1371/journal.pbio.1000344>.
- Tullin, S., Sams, A., Br, t J., Dahl, K., Gong, W., Jeppesen, C.B., Krogh, T.N., Olsen, G.S., Liu, Y., Pedersen, A.A., Petersen, J.M., Rolin, B., Wahlund, P.O., Kalthoff, C., 2012. Recombinant adiponectin does not lower plasma glucose in animal models of type 2 diabetes. *PLoS One* 7.
- Varga, O.E., Zsíros, N., Olsson, I.A.S., 2015. Estimating the predictive validity of diabetic animal models in rosiglitazone studies: predictive validity of diabetic animal models. *Obes. Rev.* 16, 498–507. <https://doi.org/10.1111/obr.12278>.
- Wang, B., Chandrasekera, P.C., Pippin, J.J., 2014. Leptin- and leptin receptor-deficient rodent models: relevance for human type 2 diabetes. *Curr. Diabetes Rev.* 10, 131–145.
- Wei, D., Tang, K., Wang, Q., Estill, J., Yao, L., Wang, X., Chen, Y., Yang, K., 2016. The use of GRADE approach in systematic reviews of animal studies: use GRADE in animal study. *J. Evid. Base Med.* 9, 98–104. <https://doi.org/10.1111/jebm.12198>.